# CHANGE-POINT IN STOCHASTIC DESIGN REGRESSION AND THE BOOTSTRAP

Emilio Seijo and Bodhisattva Sen
Columbia University

**Abstract**

In this paper we study the consistency of different bootstrap procedures for constructing confidence intervals (CIs) for the unique jump discontinuity (change-point) in an otherwise smooth regression function in a stochastic design setting. This problem exhibits nonstandard asymptotics and we argue that the standard bootstrap procedures in regression fail to provide valid confidence intervals for the change-point. We propose a version of smoothed bootstrap, illustrate its remarkable finite sample performance in our simulation study, and prove the consistency of the procedure. The $m$ out of $n$ bootstrap procedure is also considered and shown to be consistent. We also provide sufficient conditions for any bootstrap procedure to be consistent in this scenario.

## 1 Introduction

Change-point models may arise when a stochastic system is subject to sudden external influences and are encountered in almost every field of science. In the simplest form the model considers a random vector $X = (Y, Z)$ satisfying the following relation:

$$Y = \alpha_0 \mathbf{1}_{Z \leq \zeta_0} + \beta_0 \mathbf{1}_{Z > \zeta_0} + \epsilon, \tag{1}$$

where $Z$ is a continuous random variable, $\alpha_0 \neq \beta_0 \in \mathbb{R}$, $\zeta_0 \in [a, b] \subset \mathbb{R}$ and $\epsilon$ is a continuous random variable, independent of $Z$ with zero expectation and finite variance $\sigma^2 > 0$. The parameter of interest is $\zeta_0$, the change-point.

Despite its simplicity, model (1) captures the inherent "non-standard" nature of the problem: The least squares estimator of the change-point $\zeta_0$ converges at a rate of $n^{-1}$ to a minimizer of a two-sided, compound Poisson process that depends crucially on the entire error distribution, the marginal density of $Z$, among other nuisance parameters; see Pons (2003), Kosorok (2008b) (Section 14.5.1, pages

1

271–277) or Koul et al. (2003). Therefore, it is not practical to use this limiting distribution to build CIs for $\zeta_0$. Bootstrap methods bypass the estimation of nuisance parameters and are generally reliable in $\sqrt{n}$-convergence problems. In this paper we investigate the performance (both theoretically and through simulation) of different bootstrap schemes in building CIs for $\zeta_0$. We hope that the analysis of the bootstrap procedures employed in this paper will help illustrate the issues that arise when the bootstrap is applied in such non-standard problems.

The problem of estimating a jump-discontinuity (change-point) in an otherwise smooth curve has been under study for at least the last forty years. More recently, it has been extensively studied in the nonparametric regression and survival analysis literature; see for instance Gijbels et al. (1999), Dempfle and Stute (2002), Pons (2003), Kosorok and Song (2007), Lan et al. (2009) and the references therein. Bootstrap techniques have also been applied in many instances in change point models. Dümbgen (1991) proposed asymptotically valid confidence regions for the change-point by inverting bootstrap tests in a one-sample problem. Hŭsková and Kirch (2008) considered bootstrap CIs for the change-point of the mean in a time series context. Kosorok and Song (2007) use a form of parametric bootstrap to estimate the distribution of the estimated change-point in a stochastic design regression model that arises in survival analysis. Gijbels et al. (2004), in a slightly different setting, suggested a bootstrap procedure for model (1), but did not give a complete proof of its validity.

Our work goes beyond those cited above as follows: We present strong theoretical and empirical evidence to suggest the *inconsistency* of the two most natural bootstrap procedures in a regression setup – the usual nonparametric bootstrap (i.e., sampling from the empirical cumulative distribution function (ECDF) of $(Y, Z)$, often also called as bootstrapping "pairs") and the "residual" bootstrap. The bootstrap estimators built by both of these methods are the smallest maximizers of certain stochastic processes. We show that these processes do not have any weak limit in probability. This fact strongly suggests not only inconsistency but also the absence of *any* weak limit for the bootstrap estimators. In addition, we prove that independent sampling from a smooth approximation to the marginal of $Z$ and the centered ECDF of the residuals, and the $m$ out of $n$ bootstrap from the ECDF of $(Y, Z)$ yield asymptotically valid CIs for $\zeta_0$. The finite sample performance of the different bootstrap methods shows the superiority of the proposed smoothed bootstrap procedure. We also develop a series of convergence results which generalize those obtained in Kosorok (2008b) to triangular arrays of ran-

2

dom vectors and can be used to validate the consistency of *any* bootstrap scheme in this setup. Moreover, in the process of achieving this we develop convergence results for stochastic processes with a three-dimensional parameter which are continuous on the first two arguments and cádlág on the third. In particular, we prove a version of the argmax continuous mapping theorem for these processes which may be of independent interest (see Section A.1.1).

Although we develop our results in the setting of (1), our conclusions have broader implications (as discussed in Section 7). They extend immediately to regression functions with parametrically specified models on either side of the change-point. The *smoothed bootstrap* procedure can also be modified to work in more general nonparametric settings. Gijbels et al. (1999) consider jump-point estimation in the more general setup of non-parametric regression and develop two-stage procedures to build CI for the change-point. In the second stage of their procedure, they localize to a neighborhood of the change-point and reduce the problem to exactly that of (1). Lan et al. (2009) consider a two-stage adaptive sampling procedure to estimate the jump discontinuity. The second stage of their method relies on an approximate CI for the change-point, and the bootstrap methods developed in this paper can be immediately used in their context.

The paper is organized in the following manner: In Section 2 we describe the problem in greater detail, introduce the bootstrap schemes and describe the appropriate notion of consistency. In Section 3, we prove a series of convergence results that generalize those obtained in Kosorok (2008b). These results will constitute the general framework under which the bootstrap schemes will be analyzed. In Section 4 we study the inconsistency of the standard bootstrap methods, including the ECDF and residual bootstraps. In Section 5 we propose two bootstrap procedures and show their consistency. We compare the finite sample performance of the different bootstrap methods through a simulation study in Section 6. Finally, in Section 7 we discuss the consequences of our analysis in more general change-point regression models. Additionally, we include an Appendix with the proofs and some necessary lemmas and results.

## 2   The problem and the bootstrap schemes

Assume that we are given an i.i.d. sequence of random vectors $\{X_n = (Y_n, Z_n)\}_{n=1}^{\infty}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ having a common distribution $\mathbb{P}$ satisfying (1) for some parameter $\theta_0 := (\alpha_0, \beta_0, \zeta_0) \in \Theta := \mathbb{R}^2 \cup [a, b]$. This is a semi-

parametric model with an Euclidean parameter $\theta_0$ and two infinite-dimensional parameters – the distributions of $Z$ and $\epsilon$. We are interested in estimating $\zeta_0$, the change-point. For technical reasons, we will also assume that $\mathbb{P}(|\epsilon|^3) < \infty$. Here, and in the remaining of the paper, we take the convention that for any probability distribution $\mu$, we will denote the expectation operator by $\mu(\cdot)$. In addition, we suppose that $Z$ has a uniformly bounded, strictly positive density $f$ (with respect to the Lebesgue measure) on $[a, b]$ such that $\inf_{|z-\zeta_0| \leq \eta} f(z) > \kappa > 0$ for some $\eta > 0$ and that $\mathbb{P}(Z < a) \wedge \mathbb{P}(Z > b) > 0$. For $\theta = (\alpha, \beta, \zeta) \in \Theta$, $x = (y, z) \in \mathbb{R}^2$ write

$$m_\theta(x) := -\left(y - \alpha \mathbf{1}_{z \leq \zeta} - \beta \mathbf{1}_{z > \zeta}\right)^2, \tag{2}$$

$\mathbb{P}_n$ for the empirical measure defined by $X_1, \ldots, X_n$,

$$M_n(\theta) := \mathbb{P}_n(m_\theta) = -\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \alpha \mathbf{1}_{Z_i \leq \zeta} + \beta \mathbf{1}_{Z_i > \zeta}\right)^2, \tag{3}$$

and $M(\theta) := \mathbb{P}(m_\theta)$. The function $M_n$ is strictly concave in its first two coordinates but càdlàg (right continuous with left limits) in the third; in fact, piecewise constant and with $n$ jumps (w.p. 1). Thus, $M_n$ has unique maximizing values of $\alpha$ and $\beta$, but an entire interval of maximizers for $\zeta$. For this reason, we define the *least squares estimator* of $\theta_0$ to be the maximizer of $M_n$ over $\Theta$ with the smallest $\zeta$, and denote it by

$$\hat{\theta}_n := (\hat{\alpha}_n, \hat{\beta}_n, \hat{\zeta}_n) = \operatorname*{sargmax}_{\theta \in \Theta} \{M_n(\theta)\},$$

where sargmax stands for the *smallest argmax*. At this point we would like to clarify what we mean by a maximizer: if $W$ is a càdlàg process on an interval $I$, a point $x \in I$ is said to be a maximizer if $W(x) \vee W(x^-) = \sup\{W(s) : s \in I\}$. In the context of our problem, $(\alpha, \beta, \zeta)$ is a maximizer of $M_n$ if $M_n(\alpha, \beta, \zeta) \vee M_n(\alpha, \beta, \zeta^-) = \sup\{M_n(\theta) : (\theta) \in \Theta\}$.

The asymptotic properties of this least squares estimator are well known. It is shown in Kosorok (2008b), pages 271–277, that $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_{\mathbf{P}}(1)$, $\sqrt{n}(\hat{\beta}_n - \beta_0) = O_{\mathbf{P}}(1)$ and $n(\hat{\zeta}_n - \zeta_0) = O_{\mathbf{P}}(1)$. It is also shown that the asymptotic distribution of $n(\hat{\zeta}_n - \zeta_0)$ is that of the smallest argmax of a two-sided compound Poisson process. However, the limiting process depends on the distribution of $\epsilon$ and the value of the density of $Z$ at $\zeta_0$. Thus, there is no straightforward way to build CIs for $\zeta_0$ using this limiting distribution. In this connection we investigate the performance of bootstrap procedures for constructing CIs for $\zeta_0$.

4

## 2.1 Bootstrap

We start with a brief review of the bootstrap. Given a sample $\mathbf{W}_n = \{W_1, W_2, \ldots, W_n\} \overset{\text{iid}}{\sim} L$ from an unknown distribution $L$, suppose that the distribution function $H_n$ of some random variable $R_n \equiv R_n(\mathbf{W}_n, L)$ is of interest; $R_n$ is usually called a *root* and it can in general be any measurable function of the data and the distribution $L$. The bootstrap method can be broken into three simple steps:

(i) Construct an estimator $\hat{L}_n$ of $L$ from $\mathbf{W}_n$.

(ii) Generate $\mathbf{W}_n^* = \{W_1^*, \ldots, W_{m_n}^*\} \overset{\text{iid}}{\sim} \hat{L}_n$ given $\mathbf{W}_n$.

(iii) Estimate $H_n$ by $\hat{H}_n$, the conditional CDF of $R_n(\mathbf{W}_n^*, \hat{L}_n)$ given $\mathbf{W}_n$.

Let $d$ denote the Prokhorov metric or any other metric metrizing weak convergence of probability measures. We say that $\hat{H}_n$ is *weakly consistent* if $d(H_n, \hat{H}_n) \overset{P}{\to} 0$; if $H_n$ has a weak limit $H$, this is equivalent to $\hat{H}_n$ converging weakly to $H$ in probability. Similarly, $\hat{H}_n$ is *strongly consistent* if $d(H_n, \hat{H}_n) \overset{a.s.}{\to} 0$.

The choice of $\hat{L}_n$ mostly considered in the literature is the ECDF. Intuitively, an $\hat{L}_n$ that mimics the essential properties (e.g., smoothness) of the underlying distribution $L$ can be expected to perform well. Despite being a good estimator in most situations, the ECDF can fail to capture some properties of L that may be crucial for the problem under consideration. This is especially true for nonstandard problems. In Section 4 we illustrate this phenomenon (the inconsistency of the ECDF bootstrap) when $n(\hat{\zeta}_n - \zeta_0)$ is the random variable (root) of interest.

We denote by $\mathfrak{X} = \sigma\left((X_n)_{n=1}^{\infty}\right)$ the $\sigma$-algebra generated by the sequence $(X_n)_{n=1}^{\infty}$ and write $\mathbf{P}_{\mathfrak{X}}(\cdot) = \mathbf{P}(\cdot|\mathfrak{X})$ and $\mathbf{E}_{\mathfrak{X}}(\cdot) = \mathbf{E}(\cdot|\mathfrak{X})$. We approximate the CDF of $\Delta_n = n(\hat{\zeta}_n - \zeta_0)$ by $\mathbf{P}_{\mathfrak{X}}(\Delta_n^* \leq x)$, the conditional distribution function of $\Delta_n^* = m_n(\zeta_n^* - \hat{\zeta}_n)$ and use this to build a CI for $\zeta_0$, where $\zeta_n^*$ is the least squares estimator of $\zeta_0$ obtained from the bootstrap sample. In the following we introduce four bootstrap schemes that arise naturally in this problem and investigate their consistency properties in Sections 4 and 5.

**Scheme 1 (ECDF bootstrap)**: Draw a bootstrap sample $(Y_{n,1}^*, Z_{n,1}^*), \ldots, (Y_{n,n}^*, Z_{n,n}^*)$ from the ECDF of $(Y_1, Z_1), \ldots, (Y_n, Z_n)$; probably the most widely used bootstrap scheme.

**Scheme 2 (Bootstrapping residuals)**: This is another widely used bootstrap procedure in regression models. We first obtain the residuals

$$\hat{\epsilon}_{n,j} := Y_j - \hat{\alpha}_n \mathbf{1}_{Z_j \leq \hat{\zeta}_n} - \hat{\beta}_n \mathbf{1}_{Z_j > \hat{\zeta}_n} \quad \text{for } j = 1, \ldots, n,$$

from the fitted model. Note that these residuals are not guaranteed to have mean 0, so we work with the centered residuals, $\hat{\epsilon}_{n,1} - \bar{\epsilon}_n, \ldots, \hat{\epsilon}_{n,n} - \bar{\epsilon}_n$, where $\bar{\epsilon}_n = \sum_{j=1}^{n} \hat{\epsilon}_{n,j}/n$. Letting $\mathbb{P}_n^{\epsilon}$ denote the empirical measure of the centered residuals, we obtain the bootstrap sample $(Y_{n,1}^*, Z_1), \ldots, (Y_{n,n}^*, Z_n)$ as:

1. Sample $\epsilon_{n,1}^*, \ldots, \epsilon_{n,n}^*$ independently from $\mathbb{P}_n^{\epsilon}$.

2. Fix the predictors $Z_j$, $j = 1, \ldots, n$, and define the bootstrapped responses at $Z_j$ as $Y_{n,j}^* = \hat{\alpha}_n \mathbf{1}_{Z_j \leq \hat{\zeta}_n} + \hat{\beta}_n \mathbf{1}_{Z_j > \hat{\zeta}_n} + \epsilon_{n,j}^*$.

**Scheme 3 (Smoothed bootstrap)**: Notice that in (1), $Z$ is assumed to have a density and it also arises in the limiting distribution of $\Delta_n$. A successful bootstrap scheme must mimic this underlying assumption, and we accomplish this in the following:

1. Choose an appropriate nonparametric smoothing procedure (e.g., kernel density estimation) to build a distribution $\hat{F}_n$ with a density $\hat{f}_n$ such that $\|\hat{F}_n - F\|_\infty \overset{a.s.}{\to} 0$ and $\hat{f}_n \to f$ uniformly on some open interval around $\zeta_0$ w.p. 1, where $f$ is the density of $Z$.

2. Get i.i.d. replicates $Z_{n,1}^*, \ldots, Z_{n,n}^*$ from $\hat{F}_n$ and sample, independently, $\epsilon_{n,1}^*, \ldots, \epsilon_{n,n}^*$ from $\mathbb{P}_n^{\epsilon}$.

3. Define $Y_{n,j}^* = \hat{\alpha}_n \mathbf{1}_{Z_{n,j}^* \leq \hat{\zeta}_n} + \hat{\beta}_n \mathbf{1}_{Z_{n,j}^* > \hat{\zeta}_n} + \epsilon_{n,j}^*$ for all $j = 1, \ldots, n$.

**Scheme 4 ($m$ out of $n$ bootstrap)**: A natural alternative to the usual nonparametric bootstrap (i.e., generating bootstrap samples from the ECDF) considered widely in non-regular problems is to use the $m$ out of $n$ bootstrap. We choose a nondecreasing sequence of natural numbers $\{m_n\}_{n=1}^{\infty}$ such that $m_n = o(n)$ and $m_n \to \infty$ and generate the bootstrap sample $(Y_{n,1}^*, Z_{n,1}^*), \ldots, (Y_{n,m_n}^*, Z_{n,m_n}^*)$ from the ECDF of $(Y_1, Z_1), \ldots, (Y_n, Z_n)$. Although there are a number of methods available for choosing the $m_n$ in applications, there is no satisfactory solution to this problem and the obtained CIs usually vary with changing $m_n$.

We will use the framework established by our convergence theorems in Section 3 to prove that schemes 3 and 4 above yield *consistent* bootstrap procedures for building CIs for $\zeta_0$. We will also give strong empirical and theoretical evidence for the *inconsistency* of schemes 1 and 2. Note that schemes 1 and 2 are the two most widely used resampling techniques in regression models (see pages 35-36 of Efron (1982);

6

also see Freedman (1981) and Wu (1986)). Thus in this change–point scenario, a typical nonstandard problem, we see that the two standard bootstrap approaches fail. The failure of the usual bootstrap methods in nonstandard situations is not new and has been investigated in the context of M-estimation problems by Bose and Chatterjee (2001) and in situations giving rise to $n^{1/3}$ asymptotics by Abrevaya and Huang (2005) and Sen et al. (2010). But the change-point problem considered in this paper is indeed quite different from the nonstandard problems considered by the above authors – one key distinction being that compound Poisson processes, as opposed to Gaussian processes, form the backbone of the asymptotic distributions of the estimators – and thus demands an independent investigation. We will also see later that the performance of scheme 3 clearly dominates that of the $m$ out of $n$ bootstrap procedure (scheme 4), the general recipe proposed in situations where the usual bootstrap does not work (see Lee and Pun (1981) for applications of the $m$ out of $n$ bootstrap procedure in some nonstandard problems). Also note that the performance of the $m$ out of $n$ bootstrap scheme crucially depends on $m$ (see e.g., Bickel et al. (1997)) and the choice of this tuning parameter is tricky in applications.

# 3   A uniform convergence result

In this section we generalize the results obtained in Kosorok (2008b), pages 271–277, to a triangular array of random variables. Consider the triangular array $\{X_{n,k} = (Y_{n,k}, Z_{n,k})\}_{1 \leq k \leq m_n}^{n \in \mathbb{N}}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$, where $(m_n)_{n=1}^{\infty}$ is a nondecreasing sequence of natural numbers such that $m_n \to \infty$. Throughout the entire paper we will always denote by $\mathbf{E}$ the expectation operator with respect to $\mathbf{P}$. Furthermore, assume that for each $n \in \mathbb{N}$, $(X_{n,1}, \ldots, X_{n,m_n})$ constitutes a random sample from an arbitrary bivariate distribution $\mathbb{Q}_n$ with $\mathbb{Q}_n(Y_{n,1}^2) < \infty$ and let $M_n(\theta) := \mathbb{Q}_n(m_\theta)$ for all $\theta \in \Theta$, where $m_\theta$ is defined in (2). Let $\mathbb{P}$ be a bivariate distribution satisfying (1). Recall that $M(\theta) := \mathbb{P}(m_\theta)$ and $\theta_0 := \operatorname{sargmax} M(\theta)$. Let $\theta_n := (\alpha_n, \beta_n, \zeta_n)$ be given by

$$\theta_n = \operatorname*{sargmax}_{\theta \in \Theta} \{\mathbb{Q}_n(m_\theta)\}.$$

Note that $\mathbb{Q}_n$ need not satisfy model (1) with $(\alpha_n, \beta_n, \zeta_n)$. The existence of $\theta_n$ is guaranteed as $\mathbb{Q}_n(m_\theta)$ is a quadratic function in $\alpha$ and $\beta$ (for a fixed $\zeta$) and bounded and cádlág as a function in $\zeta$. For each $n$, let $\mathbb{P}_n^*$ be the empirical measure produced by the random sample $(X_{n,1}, \ldots, X_{n,m_n})$, and define the least squares

estimator $\theta_n^* = (\alpha_n^*, \beta_n^*, \zeta_n^*) \in \Theta$ to be the smallest argmax of $M_n^*(\theta) := \mathbb{P}_n^*(m_\theta)$. If $Q$ is a signed Borel measure on $\mathbb{R}^2$ and $\mathscr{F}$ is a class of (possibly) complex-valued functions defined on $\mathbb{R}^2$, write $\|Q\|_{\mathscr{F}} := \sup\{|Q(f)| : f \in \mathscr{F}\}$. If $g : K \subset \mathbb{R}^3 \to \mathbb{R}$ is a bounded function, write $\|g\|_K := \sup_{x \in K} |g(x)|$. Also, for $(z, y) \in \mathbb{R}^2$ and $n \in \mathbb{N}$ we write

$$\tilde{\epsilon}_n := \tilde{\epsilon}_n(z, y) = y - \alpha_n \mathbf{1}_{z \leq \zeta_n} - \beta_n \mathbf{1}_{z > \zeta_n}. \tag{4}$$

Let $M > 0$ be such that $|\alpha_n| \leq M$ for all $n$. We define the following three classes of functions from $\mathbb{R}^2$ into $\mathbb{R}$:

$$\begin{aligned}
\mathcal{F} &:= \{\mathbf{1}_I(z) : I \subset \mathbb{R} \text{ is an interval}\}, \\
\mathcal{G} &:= \{yf(z) : f \in \mathcal{F}\} \cup \{|y + \alpha|f(z) : f \in \mathcal{F}, |\alpha| \leq M\}, \\
\mathcal{H} &:= \{y^2 f(z) : f \in \mathcal{F}\}.
\end{aligned}$$

In what follows, we will derive conditions on the distributions $\mathbb{Q}_n$ that will guarantee consistency and weak convergence of $\theta_n^*$.

## 3.1 Consistency and the rate of convergence

We provide first a consistency result for the least squares estimator, whose proof we include in the Appendix (see Section A.2.1). To this end, we consider the following set of assumptions:

(I) $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \to 0$,

(II) $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{G}} \to 0$,

(III) $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{H}} \to 0$,

(IV) $\theta_n \to \theta_0$.

**Proposition 3.1** *Assume that (I)-(IV) hold. Then, $\theta_n^* \xrightarrow{\mathbf{P}} \theta_0$.*

To guarantee the right rate of convergence, we need to assume stronger regularity conditions. In addition to those of Proposition 3.1, we require the following:

(V) There are $\eta, \rho, L > 0$ with the property that for any $\delta \in (0, \eta)$, there is $N > 0$

such that the following inequalities hold for any $n \geq N$:

$$\inf_{\frac{1}{\sqrt{m_n}} \leq |\zeta - \zeta_n| < \delta^2} \left\{ \frac{1}{|\zeta - \zeta_n|} \mathbb{Q}_n(\mathbf{1}_{\zeta \wedge \zeta_n < Z \leq \zeta \vee \zeta_n}) \right\} > \rho, \qquad (5)$$

$$\sup_{|\zeta - \zeta_n| < \delta^2} \left\{ |\mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{\zeta \wedge \zeta_n < Z \leq \zeta \vee \zeta_n})| \right\} \leq \frac{L\delta}{\sqrt{m_n}}, \qquad (6)$$

$$\sup_{|\zeta - \zeta_n| < \delta^2} \left\{ |\mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{Z \leq \zeta \wedge \zeta_n})| + |\mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta \vee \zeta_n})| \right\} \leq \frac{L}{\sqrt{m_n}}. \qquad (7)$$

We would like to point out some facts about (V). It must be noted that (6) and (7) automatically hold in the case where $Z$ and $\tilde{\epsilon}_n$ are independent under $\mathbb{Q}_n$ with $\mathbb{Q}_n(\tilde{\epsilon}_n) = 0$. Also, (5) is easily seen to hold when the $Z$'s, under $\mathbb{Q}_n$, have densities $f_n$ converging uniformly to $f$ in some neighborhood of $\zeta_0$, where $f$ is the density of $Z$ under $\mathbb{P}$; by a consequence of the classical mean value theorem of calculus. With the aid of these conditions, Proposition 3.1 and Theorem 3.4.1, page 322, of Van der Vaart and Wellner (1996) we can now state and prove (see Section A.2.2) the rate of convergence result.

**Proposition 3.2** *Assume that (I)-(V) hold. Then* $\sqrt{m_n}(\alpha_n^* - \alpha_n) = O_{\mathbf{P}}(1)$, $\sqrt{m_n}(\beta_n^* - \beta_n) = O_{\mathbf{P}}(1)$ *and* $m_n(\zeta_n^* - \zeta_n) = O_{\mathbf{P}}(1)$.

Propositions 3.1 and 3.2 provide sufficient conditions on the measures $\mathbb{Q}_n$, the distribution of each element in the $n$th row of the triangular array, to achieve the same rate of convergence as the original least squares estimators. We would like to highlight that we are not assuming that each $\mathbb{Q}_n$ satisfy the model (1) with $(\alpha_n, \beta_n, \zeta_n)$; all we need is that $\mathbb{Q}_n$ and $\theta_n$ approach $\mathbb{P}$ and $\theta_0$ respectively, in a suitable manner.

## 3.2 Weak Convergence and asymptotic distribution

We start with some additional set of assumptions:

(VI) For any function $\psi : \mathbb{R} \to \mathbb{C}$ which is either of the form $\psi(x) = e^{i\xi x}$ for some $\xi \in \mathbb{R}$ or defined by $\psi(x) = |x|^p$ for $p = 1, 2$, we have:

$$m_n \mathbb{Q}_n \left( \psi(\tilde{\epsilon}_n) \mathbf{1}_{\zeta_n - \frac{\delta}{m_n} < Z \leq \zeta_n + \frac{\eta}{m_n}} \right) \to f(\zeta_0)(\delta + \eta) \mathbb{P}(\psi(\epsilon)) \quad \forall \, \eta, \delta > 0.$$

(VII) $\sqrt{m_n} \mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{Z \leq \zeta_n}) \to 0$ and $\sqrt{m_n} \mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta_n}) \to 0$.

(VIII) $\overline{\lim}_{n \to \infty} \mathbb{Q}_n(|\tilde{\epsilon}_n|^3) < \infty$.

Observe that condition (VI) implies, for all $\eta, \delta > 0$, and $p = 1, 2$,

$$\sqrt{m_n}\mathbb{Q}_n \left( |\tilde{\epsilon}_n|^p \mathbf{1}_{\zeta_n - \frac{\delta}{m_n} < Z \leq \zeta_n + \frac{\eta}{m_n}} \right) \quad \to \quad 0, \tag{8}$$

$$\sqrt{m_n}\mathbb{Q}_n \left( \mathbf{1}_{\zeta_n - \frac{\delta}{m_n} < Z \leq \zeta_n + \frac{\eta}{m_n}} \right) \quad \to \quad 0. \tag{9}$$

For $h = (h_1, h_2, h_3) \in \mathbb{R}^3$, let $\vartheta_{n,h} := \theta_n + \left( \frac{h_1}{\sqrt{m_n}}, \frac{h_2}{\sqrt{m_n}}, \frac{h_3}{m_n} \right)$ and

$$\hat{E}_n(h) := m_n \mathbb{P}_n^* \left[ m_{\vartheta_{n,h}} - m_{\theta_n} \right].$$

We will argue that

$$h_n^* := \operatorname*{sargmin}_{h \in \mathbb{R}^3} \hat{E}_n(h) = (\sqrt{m_n}(\alpha_n^* - \alpha_n), \sqrt{m_n}(\beta_n^* - \beta_n), m_n(\zeta_n^* - \zeta_n))$$

converges in distribution to the smallest argmax of some process involving two independent normal random variables and a two-sided, compound Poisson process (independent of the normal variables).

We derive the asymptotic distribution of the process $\hat{E}_n$ and then apply continuous mapping techniques to obtain the limiting distribution of $h_n^*$. We consider these stochastic processes as random elements in the space $\mathcal{D}_K$, for a given compact rectangle $K \subset \mathbb{R}^3$, of all functions $W : K \to \mathbb{R}$ having "quadrant limits" (as defined in Neuhaus (1971)), being continuous from above (again, in the terminology of Neuhaus (1971)) and such that $W(\cdot, \cdot, \zeta)$ is continuous for all $\zeta$ and $W(\alpha, \beta, \cdot)$ is càdlàg (right continuous having left limits) for all $(\alpha, \beta)$. Write $\mathcal{D} = \mathcal{D}_{\mathbb{R}^3}$. For any compact interval $I \subset \mathbb{R}$ let

$$\Lambda_I = \{\lambda : I \to I \mid \lambda \text{ is strictly increasing, surjective and continuous}\}$$

and write

$$\|\lambda\| := \sup_{s \neq t \in I} \left| \log \frac{\lambda(s) - \lambda(t)}{s - t} \right|.$$

Then, for any set of the form $K = A \times I$ with $A \subset \mathbb{R}^2$ define the Skorohod topology as the topology given by the metric

$$d_K(\Psi, \Gamma) := \inf_{\lambda \in \Lambda_I} \left\{ \sup_{(\alpha, \beta, \zeta) \in K} \{ |\Psi(\alpha, \beta, \zeta) - \Gamma(\alpha, \beta, \lambda(\zeta))| \} + \|\lambda\| \right\}$$

for $\Gamma, \Psi \in \mathcal{D}_K$. Endowed with this metric, $\mathcal{D}_K$ becomes a Polish space (it is a closed subspace of the Polish spaces $D_k$ defined in Neuhaus (1971)) and thus the existence of conditional probability distributions for its random elements is ensured

10

(see Dudley (2002), Theorem 10.2.2 page 345). Also, let $\tilde{\mathcal{D}}_I$, $I \subset \mathbb{R}$, denote the space of real valued càdlàg functions on $I$. We refer the reader to Section A.1 for some results about the Skorohod space.

We express the process $\hat{E}_n$ as the sum of the four terms $\hat{A}_n$, $\hat{B}_n$, $\hat{C}_n$ and $\hat{D}_n$ where

$$
\begin{aligned}
\hat{A}_n(h_1, h_3) &:= 2h_1\sqrt{m_n}\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z \le \zeta_n \wedge \left(\zeta_n + \frac{h_3}{m_n}\right)}\right) - h_1^2 \mathbb{P}_n^*\left(\mathbf{1}_{Z \le \zeta_n \wedge \left(\zeta_n + \frac{h_3}{m_n}\right)}\right), \\
\hat{B}_n(h_2, h_3) &:= 2h_2\sqrt{m_n}\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta_n \vee \left(\zeta_n + \frac{h_3}{m_n}\right)}\right) - h_2^2 \mathbb{P}_n^*\left(\mathbf{1}_{Z > \zeta_n \vee \left(\zeta_n + \frac{h_3}{m_n}\right)}\right), \\
\hat{C}_n(h_2, h_3) &:= -2m_n\left(\alpha_n - \beta_n + \frac{h_2}{\sqrt{m_n}}\right)\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n + \frac{h_3}{m_n} < Z \le \zeta_n}\right) \\
&\qquad - m_n\left(\alpha_n - \beta_n + \frac{h_2}{\sqrt{m_n}}\right)^2 \mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n + \frac{h_3}{m_n} < Z \le \zeta_n}\right), \\
\hat{D}_n(h_1, h_3) &:= -2m_n\left(\beta_n - \alpha_n + \frac{h_1}{\sqrt{m_n}}\right)\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n < Z \le \zeta_n + \frac{h_3}{m_n}}\right) \\
&\qquad - m_n\left(\beta_n - \alpha_n + \frac{h_1}{\sqrt{m_n}}\right)^2 \mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n < Z \le \zeta_n + \frac{h_3}{m_n}}\right).
\end{aligned}
$$

We define another process $E_n^* := A_n^* + B_n^* + C_n^* + D_n^*$ where

$$
\begin{aligned}
A_n^*(h_1) &:= 2h_1\sqrt{m_n}\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z \le \zeta_n}\right) - h_1^2 \mathbb{P}_n^*\left(\mathbf{1}_{Z \le \zeta_n}\right), \\
B_n^*(h_2) &:= 2h_2\sqrt{m_n}\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta_n}\right) - h_2^2 \mathbb{P}_n^*\left(\mathbf{1}_{Z > \zeta_n}\right), \\
C_n^*(h_3) &:= -2m_n(\alpha_n - \beta_n)\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n + \frac{h_3}{m_n} < Z \le \zeta_n}\right) \\
&\qquad - m_n(\alpha_n - \beta_n)^2 \mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n + \frac{h_3}{m_n} < Z \le \zeta_n}\right), \\
D_n^*(h_3) &:= -2m_n(\beta_n - \alpha_n)\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n < Z \le \zeta_n + \frac{h_3}{m_n}}\right) \\
&\qquad - m_n(\beta_n - \alpha_n)^2 \mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n < Z \le \zeta_n + \frac{h_3}{m_n}}\right).
\end{aligned}
$$

We work with $E_n^*$ instead of $\hat{E}_n$ as their difference approaches uniformly to 0 in probability, as shown in the next lemma (proved in Section A.2.3), and the asymptotic distribution of $E_n^*$ is easier to derive.

**Lemma 3.1** *Let $K \subset \mathbb{R}^3$ be a compact rectangle. If conditions (I)-(IV) and (8) and (9) hold, then*

$$
\left\| E_n^* - \hat{E}_n \right\|_K \xrightarrow{\mathbf{P}} 0.
$$

*Therefore, $E_n^* - \hat{E}_n \xrightarrow{\mathbf{P}} 0$ as random elements of $\mathcal{D}_K$. In particular, this result is true under conditions (I)-(IV) and (VI).*

11

As a first step to finding the asymptotic distribution of $(E_n^*)_{n=1}^\infty$, we show that the random sequence is tight in the Skorohod space $\mathcal{D}_K$ for any compact rectangle $K \subset \mathbb{R}^3$. The proof of the next result is given in Section A.2.4.

**Lemma 3.2** *Let $I \subset \mathbb{R}$ be a compact interval and assume that conditions (I)-(VIII) hold. Then, the sequence of $\mathbb{R}^6$-valued processes*

$$
\Xi_n(t) := \begin{pmatrix}
\sqrt{m_n}\mathbb{P}_n^*(\tilde{\epsilon}_n \mathbf{1}_{Z \leq \zeta_n}) \\
\sqrt{m_n}\mathbb{P}_n^*(\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta_n}) \\
m_n \mathbb{P}_n^*(\mathbf{1}_{\zeta_n + \frac{t}{m_n} < Z \leq \zeta_n}) \\
m_n \mathbb{P}_n^*(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n + \frac{t}{m_n} < Z \leq \zeta_n}) \\
m_n \mathbb{P}_n^*(\mathbf{1}_{\zeta_n < Z \leq \zeta_n + \frac{t}{m_n}}) \\
m_n \mathbb{P}_n^*(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n < Z \leq \zeta_n + \frac{t}{m_n}})
\end{pmatrix}
\tag{10}
$$

*is uniformly tight in $\mathbb{R}^2 \times \tilde{\mathcal{D}}_I^4$. Also, if $K \subset \mathbb{R}^3$ is a compact rectangle, the sequence $(E_n^*)_{n=1}^\infty$ is uniformly tight in $\mathcal{D}_K$.*

It now suffices to show convergence of the finite-dimensional distributions of the processes $E_n^*$ to the finite dimensional distributions of some process $E^* \in \mathcal{D}_K$ to conclude that $E_n^*$ converges weakly to $E^*$ (and thus $\hat{E}_n$ too). With this objective in mind, we make the following definitions: Let $\mathbf{Z}_1 \sim \mathbf{N}(0, \sigma^2 \mathbb{P}(Z \leq \zeta_0))$ and $\mathbf{Z}_2 \sim \mathbf{N}(0, \sigma^2 \mathbb{P}(Z > \zeta_0))$ be two independent normal random variables; $\nu_1$ and $\nu_2$ be, respectively, left-continuous and right-continuous, homogeneous Poisson processes with rate $f(\zeta_0) > 0$; $\mathbf{u} = (u_n)_{n=1}^\infty$ and $\mathbf{v} = (v_n)_{n=1}^\infty$ two sequences of i.i.d. random variables having the same distribution as $\epsilon$ under $\mathbb{P}$. Assume, in addition, that $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\nu_1$, $\nu_2$, $\mathbf{v}$ and $\mathbf{u}$ are all mutually independent. Then, define the process $\Xi = (\Xi^{(1)}, \ldots, \Xi^{(6)})'$ as

$$
\Xi(t) := \begin{pmatrix}
\mathbf{Z}_1 \\
\mathbf{Z}_2 \\
\nu_1(-t)\mathbf{1}_{t<0} \\
\sum_{0 < j \leq \nu_1(-t)} v_j \mathbf{1}_{t<0} \\
\nu_2(t)\mathbf{1}_{t \geq 0} \\
\sum_{0 < j \leq \nu_2(t)} u_j \mathbf{1}_{t \geq 0}
\end{pmatrix}
\tag{11}
$$

and let $E^*$ be given by

$$
\begin{aligned}
E^*(h) \quad := \quad & 2h_1 \Xi^{(1)}(h_3) - h_1^2 \mathbb{P}(Z \leq \zeta_0) + 2h_2 \Xi^{(2)}(h_3) - h_2^2 \mathbb{P}(Z > \zeta_0) \\
& + 2(\beta_0 - \alpha_0)\Xi^{(4)}(h_3) - (\alpha_0 - \beta_0)^2 \Xi^{(3)}(h_3) \\
& + 2(\alpha_0 - \beta_0)\Xi^{(6)}(h_3) - (\alpha_0 - \beta_0)^2 \Xi^{(5)}(h_3)
\end{aligned}
\tag{12}
$$

for $h = (h_1, h_2, h_3) \in \mathbb{R}^3$.

We will now prove weak convergence of the sequence of processes $(\hat{E}_n)_{n=1}^{\infty}$ to $E^*$, and then use a continuous mapping theorem for the smallest argmax functional (see Lemma A.3) to obtain weak convergence of $h_n^* := \text{sargmax } \hat{E}_n(h)$. The application of Lemma A.3 requires the weak convergence of processes $(\hat{E}_n)_{n=1}^{\infty}$ to $E^*$ and also the weak convergence of their associated jump processes. Let $\mathcal{S}$ be the class of all piecewise constant, cádlág functions $\tilde{\psi} : \mathbb{R} \to \mathbb{R}$ that are continuous on the integers with $\tilde{\psi}(0) = 0$; $\tilde{\psi}$ has jumps of size 1, and $\tilde{\psi}(-t)$ and $\tilde{\psi}(t)$ are nondecreasing on $(0, \infty)$. For an interval $I$ containing 0 in its interior, we write $\mathcal{S}_I = \{f \mid_I : f \in \mathcal{S}\}$. Define the $\mathcal{S}$–valued (pure jump) processes $\hat{J}_n$, $J_n^*$ and $J^*$ as

$$
\begin{aligned}
J_n^*(t) = \hat{J}_n(t) &:= m_n \mathbb{P}_n^*(\mathbf{1}_{\zeta_n + \frac{t}{m_n} < Z \leq \zeta_n}) + m_n \mathbb{P}_n^*(\mathbf{1}_{\zeta_n < Z \leq \zeta_n + \frac{t}{m_n}}), \\
J^*(t) &:= \nu_1(-t)\mathbf{1}_{t<0} + \nu_2(t)\mathbf{1}_{t\geq 0}.
\end{aligned}
$$

**Lemma 3.3** *Let $I \subset \mathbb{R}$ be a compact interval and $K = A \times B \times I \subset \mathbb{R}^3$ a compact rectangle. If (I)-(VIII) hold, we have*

*(i) $\Xi_n \rightsquigarrow \Xi$ in $\mathbb{R}^2 \times \tilde{\mathcal{D}}_I^4$,*

*(ii) $(E_n^*, J_n^*) \rightsquigarrow (E^*, J^*)$ in $\mathcal{D}_K \times \mathcal{S}_I$,*

*(iii) $(\hat{E}_n, \hat{J}_n) \rightsquigarrow (E^*, J^*)$ in $\mathcal{D}_K \times \mathcal{S}_I$,*

*where $\rightsquigarrow$ denotes weak convergence.*

For a proof of the convergence result, see Section A.2.5.

To apply the argmax continuous mapping theorem we first show that the the smallest argmax of $E^*$ is well defined. The proof of the next lemma is provided in Section A.2.6.

**Lemma 3.4** *Consider the process $E^*$ defined in (12). Then, for almost every sample path of $E^*$, $\phi^* = (\phi_1^*, \phi_2^*, \phi_3^*) := \underset{h \in \mathbb{R}^3}{\text{sargmax}}\{E^*(h)\}$ is well-defined. Moreover, $\phi_1^*$, $\phi_2^*$ and $\phi_3^*$ are independent; and $\phi_1^*$ and $\phi_2^*$ are distributed as normal random variables with mean 0 and variances $\sigma^2/\mathbb{P}(Z \leq \zeta_0)$ and $\sigma^2/\mathbb{P}(Z > \zeta_0)$, respectively.*

We now state the distributional convergence result for the sequence of least squares estimator $\theta_n^*$. For a proof, we refer the reader to Section A.2.7.

**Proposition 3.3** *With the notation of Lemma 3.4, if conditions (I)-(VIII) hold, then*

$$h_n^* = \begin{pmatrix} \sqrt{m_n}(\alpha_n^* - \alpha_n) \\ \sqrt{m_n}(\beta_n^* - \beta_n) \\ m_n(\zeta_n^* - \zeta_n) \end{pmatrix} \rightsquigarrow \underset{h \in \mathbb{R}^3}{sargmax}\{E^*(h)\}.$$

If we take $\mathbb{Q}_n = \mathbb{P}$ and $m_n = n \ \forall n \in \mathbb{N}$, it is easily seen that $\theta_n = \theta_0$ and conditions (I)-(VIII) hold. Hence, we immediately get the following corollary.

**Corollary 3.1 (Asymptotic distribution of the least squares estimators)** *For the least squares estimators $(\hat{\alpha}_n, \hat{\beta}_n, \hat{\zeta}_n)$ based on an i.i.d. sequence $(X_n)_{n=1}^\infty$ satisfying (1), we have*

$$(\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), n(\hat{\zeta}_n - \zeta_0))' \rightsquigarrow \underset{h \in \mathbb{R}^3}{sargmax}\{E^*(h)\}.$$

# 4 Inconsistency of the bootstrap

In this section we argue the inconsistency of the two most common bootstrap procedures in regression: the ECDF bootstrap (scheme 1) and the residual bootstrap (scheme 2). Recall the notation and definitions in the beginning of Section 2. In particular, note that we have i.i.d. random vectors $\{X_n = (Y_n, Z_n)\}_{n=1}^\infty$ from (1) with parameter $\theta_0$ defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and let $\mathbb{P}_n$ be the empirical distribution of the first $n$ data points. We start by stating two results that will be used in the sequel. We first show that the least squares estimator $\hat{\theta}_n$ of $\theta_0$ is strongly consistent. This is an improvement of the result obtained in Kosorok (2008b) and we refer the reader to Section A.2.8 for a complete proof. The proof of the second lemma can be found in Section A.2.9.

**Lemma 4.1** *Let $K \subset \Theta$ be any compact rectangle. Then,*

(i) $\|M_n - M\|_K \xrightarrow{a.s.} 0$,

(ii) $M_n \xrightarrow{a.s.} M$ in $\mathcal{D}_K$,

(iii) $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$.

**Lemma 4.2** *Let $K \subset \mathbb{R}$ be a compact interval and $(m_n)_{n=1}^\infty$ be an increasing sequence of natural numbers such that $m_n \to \infty$ and $m_n = O(n)$. Then,*

(i) $m_n^\gamma \left\| \mathbb{P}_n(\hat{\zeta}_n + \frac{(\cdot)}{m_n} < Z \le \hat{\zeta}_n) \right\|_K \xrightarrow{\mathbf{P}} 0$ *for any $\gamma < 1$, and*

14

*(ii)* $m_n^\gamma \left\| \mathbb{P}_n \left( |\tilde{\epsilon}_n|^p \mathbf{1}_{\hat{\zeta}_n + \frac{(\cdot)}{m_n} < Z \le \hat{\zeta}_n} \right) \right\|_K \xrightarrow{\mathbf{P}} 0$ *for any* $\gamma < 1$, *and* $p = 1,2$.

*These statements are still true if* $\mathbf{1}_{\hat{\zeta}_n + \frac{(\cdot)}{m_n} < Z \le \hat{\zeta}_n}$ *is replaced by* $\mathbf{1}_{\hat{\zeta}_n < Z \le \hat{\zeta}_n + \frac{(\cdot)}{m_n}}$.

We introduce some notation. Let $(\mathtt{X}, d)$ be a metric space and consider the $\mathtt{X}$-valued random elements $V$ and $(V_n)_{n=1}^\infty$ defined on $(\Omega, \mathcal{A}, \mathbf{P})$. We say that $V_n$ converges conditionally in probability to $V$, almost surely, and write $V_n \xrightarrow[a.s.]{\mathbf{P}_{\mathfrak{X}}} V$, if

$$\mathbf{P}_{\mathfrak{X}}(d(V_n, V) > \epsilon) \xrightarrow{a.s.} 0 \quad \forall \, \epsilon > 0. \tag{13}$$

Similarly, we write $V_n \xrightarrow[\mathbf{P}]{\mathbf{P}_{\mathfrak{X}}} V$ and say that $V_n$ converges conditionally in probability to $V$, in probability, if the left–hand side of (13) converges in probability to 0.

## 4.1  Scheme 1 (Bootstrapping from ECDF)

Consider the notation and definitions of Section 2.1. To translate this scheme into the framework of Propositions 3.1, 3.2 and 3.3, we set $m_n = n$, $\mathbb{Q}_n = \mathbb{P}_n$ and consider the triangular array $\left\{ X_{n,k}^* = (Y_{n,k}^*, Z_{n,k}^*) \right\}_{1 \le k \le n}^{n \in \mathbb{N}}$. Moreover, from Lemma 4.1 we know that $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$, so we can also take $\theta_n = \hat{\theta}_n$. We first prove that the bootstrapped estimators converge conditionally in probability to the true value of the parameters, almost surely.

**Proposition 4.1** *For the ECDF bootstrap, we have* $\theta_n^* \xrightarrow[a.s.]{\mathbf{P}_{\mathfrak{X}}} \theta_0$.

**Proof:**  Since $Y$ has a second moment under $\mathbb{P}$, it is straightforward to see that $\mathcal{F}, \mathcal{G}$ and $\mathcal{H}$ are VC-subgraph classes with integrable envelopes 1, $|Y| + M$ and $Y^2$, respectively. It follows that all these classes are Glivenko–Cantelli and therefore conditions (I)-(III) hold w.p. 1. Also, note that, from Lemma 4.1 *(iii)* condition (IV) holds a.s. The result then follows from Proposition 3.1.  □

Let $\mathbb{P}_n^*$ be the ECDF of $X_{n,1}^*, \dots, X_{n,n}^*$ and recall the definition of the processes $\hat{A}_n, \hat{B}_n, \hat{C}_n, \hat{D}_n, \hat{E}_n, A_n^*, B_n^*, C_n^*, D_n^*$ and $E_n^*$. We then have the following result.

**Lemma 4.3** *Let* $K \subset \mathbb{R}^3$ *be any compact rectangle. Then*

$$\hat{E}_n - E_n^* \xrightarrow[\mathbf{P}]{\mathbf{P}_{\mathfrak{X}}} 0 \ \text{in} \ \mathcal{D}_K.$$

**Proof:**  We already know that conditions (I)-(IV) hold w.p. 1 under this bootstrap scheme. But Lemma 4.2 implies that (8) and (9) hold in probability. Hence, this

15

result follows by arguing through subsequences and applying Lemma 3.1. $\qquad\square$

It is evident that condition (VI) doesn't hold in this situation as we know that

$$n\mathbb{P}_n(\zeta_0 - \frac{\eta}{n} < Z \leq \zeta_0 + \frac{\delta}{n}) \rightsquigarrow \mathrm{Poisson}\big(f(\zeta_0)(\delta + \eta)\big). \tag{14}$$

Hence, we cannot use Proposition 3.3 to derive the limit behavior of $h_n^*$.

We will now argue that $E_n^*$, and therefore $\hat{E}_n$, *does not have any weak limit* in probability. This statement should be thought in terms of the Prokhorov metric (or any other metric metrizing weak convergence on $\mathcal{D}_K$). If we denote by $\rho_K$ the Prokhorov metric on the space of probability measures on $\mathcal{D}_K$ and by $\mu_n$ the conditional distribution of $E_n^*$ given $\mathfrak{X}$, to say that $(E_n^*)_{n=1}^\infty$ has no weak limit in probability means that there is no probability measure $\mu$ defined on $\mathcal{D}_K$ such that $\rho_K(\mu_n, \mu) \xrightarrow{\mathbf{P}} 0$.

The following lemma (proved in Section A.2.10) will help us show that the (conditional) characteristic functions corresponding to the finite dimensional distributions of $E_n^*$ fail to have a limit in probability, which would, in particular, imply that $E_n^*$ does not have a weak limit in probability.

**Lemma 4.4** *The following statements hold:*

(i) *For any two real numbers $s < t$, $\left\{n\mathbb{P}_n(\zeta_0 + \frac{s}{n} < Z \leq \zeta_0 + \frac{t}{n})\right\}_{n=1}^\infty$ does not converge in probability.*

(ii) *There is $h_* > 0$ such that for any $h \geq h_*$, the sequences $\left\{n\mathbb{P}_n(\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h}{n})\right\}_{n=1}^\infty$ and $\left\{n\mathbb{P}_n(\hat{\zeta}_n - \frac{h}{n} < Z \leq \hat{\zeta}_n)\right\}_{n=1}^\infty$ do not converge in probability.*

(iii) *For any two real numbers $s < t$ and any measurable function $\phi : \mathbb{R} \to \mathbb{R}$, $\left\{n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\zeta_0 + \frac{s}{n} < Z \leq \zeta_0 + \frac{t}{n}})\right\}_{n=1}^\infty$ does not converge in probability.*

(iv) *Let $\phi$ be a measurable function which is either nonnegative or nonpositive and such that $\phi(\epsilon + \alpha_0)$ and $\phi(\epsilon + \beta_0)$ are nonconstant random variables with finite second moment. Then, there is $h_* > 0$ such that for any $h \geq h_*$ $\left\{n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h}{n}})\right\}_{n=1}^\infty$ and $\left\{n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\hat{\zeta}_n - \frac{h}{n} < Z \leq \hat{\zeta}_n})\right\}_{n=1}^\infty$ do not converge in probability.*

With the aid of Lemma 4.4 we are now able to state our main result.

**Lemma 4.5** *There is a compact rectangle $K \subset \mathbb{R}^3$ such that neither $\hat{E}_n$ nor $E_n^*$ has a weak limit in probability in $\mathcal{D}_K$.*

**Proof:** Since Lemma 4.3 and Slutsky's lemma show that $\hat{E}_n$ has a weak limit in probability if and only if $E_n^*$ has a weak limit in probability, it suffices to argue that the statement is true for $E_n^*$. To prove this, it is enough to show that there is some $h_3$ such that $E_n^*(0, 0, h_3)$ does not converge in distribution. Pick $h_3 > 0$ and observe that

$$E_n^*(0, 0, h_3) = (\hat{\alpha}_n - \hat{\beta}_n) \left( n\mathbb{P}_n^* \left[ (2\tilde{\epsilon}_n - \hat{\alpha}_n + \hat{\beta}_n)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right] \right).$$

Since $\hat{\alpha}_n - \hat{\beta}_n \xrightarrow{a.s.} \alpha_0 - \beta_0 \neq 0$ we see that $E_n^*(0, 0, h_3)$ will converge weakly in probability if and only if $\Lambda_n := n\mathbb{P}_n^* \left[ (2\tilde{\epsilon}_n - \hat{\alpha}_n + \hat{\beta}_n)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right]$ converges weakly in probability.

The conditional characteristic function of $\Lambda_n$ is given by

$$\mathbf{E}_{\mathfrak{X}} \left( e^{i\xi\Lambda_n} \right) = \left( 1 + \frac{1}{n}n\mathbb{P}_n \left( (e^{i\xi(2\tilde{\epsilon}_n + \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right) \right)^n, \tag{15}$$

which converges in probability if and only if so does

$$n\mathbb{P}_n \left( (e^{i\xi(2\tilde{\epsilon}_n + \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right).$$

. But note that

$$n\mathbb{P}_n \left( (e^{i\xi(2\tilde{\epsilon}_n + \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right) = n\mathbb{P}_n \left( (e^{i\xi(2Y - \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right).$$

It is easily seen that (14) and the fact that $n(\hat{\zeta}_n - \zeta_0) = O_{\mathbf{P}}(1)$ imply that

$$n\mathbb{P}_n \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right) = O_{\mathbf{P}}(1).$$

Hence,

$$\left| n\mathbb{P}_n \left( (e^{i\xi(2Y - \hat{\beta}_n - \hat{\alpha}_n)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right) - n\mathbb{P}_n \left( (e^{i\xi(2Y - \beta_0 - \alpha_0)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right) \right|$$

$$\leq n\mathbb{P}_n \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right) (|\hat{\alpha}_n - \alpha_0| + |\hat{\beta}_n - \beta_0|)|\xi| \xrightarrow{\mathbf{P}} 0.$$

It follows that $\mathbf{E}_{\mathfrak{X}} \left( e^{i\xi\Lambda_n} \right)$ has a limit in probability if and only if

$$n\mathbb{P}_n \left( (e^{i\xi(2Y - \beta_0 - \alpha_0)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right)$$

has a limit in probability. But a necessary condition for the latter to happen is that its real part,

$$n\mathbb{P}_n \left( \text{Re}(e^{i\xi(2Y - \beta_0 - \alpha_0)} - 1)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h_3}{n}} \right)$$

17

converges in probability. Since $\mathrm{Re}(e^{i\xi(2Y-\beta_0-\alpha_0)}-1)\le 0$ we can conclude from (iv) of Lemma 4.4 that $n\mathbb{P}_n\left(\mathrm{Re}(e^{i\xi(2Y-\beta_0-\alpha_0)}-1)\mathbf{1}_{\hat\zeta_n<Z\le\hat\zeta_n+\frac{h_3}{n}}\right)$ does not converge in probability for all $h_3\ge h_*$ for some $h_*>0$ large enough. This in turn implies that, for all $h_3\ge h_*$, the conditional characteristic function in (15) does not converge in probability and hence $E_n^*(0,0,h_3)$ has no weak limit in probability.

Hence, if $K$ is any compact rectangle containing $(0,0,h_*)$ the finite dimensional dimensional distributions of $E_n^*$ on $K$ do not have a weak limit in probability. Therefore, $E_n^*$ does not have a weak limit in probability on $\mathcal{D}_K$. $\qquad\square$

Note that

$$\left(\sqrt{n}(\alpha_n^*-\hat\alpha_n),\sqrt{n}(\beta_n^*-\hat\beta_n),n(\zeta_n^*-\hat\zeta_n)\right)=\operatorname*{sargmax}_{h\in\mathbb{R}^3}\left\{\hat E_n(h)\right\}.$$

Thus, the fact that the sequence $(\hat E_n)_{n=1}^\infty$ doesn't have a weak limit in probability makes the existence of a weak limit in probability for $n(\zeta_n^*-\hat\zeta_n)$ very unlikely. However, we do not have the a rigorous mathematical proof this statement. The main difficulty in such a proof is that the argmax functional is non-linear and that $\hat E_n$ depends on $h_3$ through indicator functions that do not converge in the limit.

**Remark:** It must be noted in this connection that the bootstrap scheme estimates the distribution of $(\sqrt{n}(\alpha_n^*-\hat\alpha_n),\sqrt{n}(\beta_n^*-\hat\beta_n))$ correctly, and in fact, valid bootstrap based inference can be conducted to obtain CIs for $\alpha_0$ and $\beta_0$. This follows from the fact that, asymptotically, the maximizers of $\hat E_n(\cdot,\cdot,h_3)$ do not depend on $h_3$ (see the expressions for $\hat A_n^*$, $\hat B_n$, $A_n^*$, $B_n^*$).

We next provide an alternative additional argument that illustrates the inconsistency of the ECDF bootstrap. Our approach is similar to that of Kosorok (2008a) and relies on the asymptotic *unconditional* behavior of

$$\tilde\Delta_n^*:=(\sqrt{n}(\alpha_n^*-\alpha_0),\sqrt{n}(\beta_n^*-\beta_0),n(\zeta_n^*-\zeta_0)).$$

For $h\in\mathbb{R}^3$, we write $\tilde\vartheta_{n,h}:=\theta_0+\left(\frac{h_1}{\sqrt{n}},\frac{h_2}{\sqrt{n}},\frac{h_3}{n}\right)$ and

$$\tilde E_n(h):=n\mathbb{P}_n^*\left[m_{\tilde\vartheta_{n,h}}-m_{\theta_0}\right]. \tag{16}$$

This corresponds to centering the objective function around $\theta_0$. As in (10), we can define the processes

$$
\tilde{\Xi}_n(t) = \begin{pmatrix} \tilde{\Xi}_n^{(1)}(t) \\ \tilde{\Xi}_n^{(2)}(t) \\ \tilde{\Xi}_n^{(3)}(t) \\ \tilde{\Xi}_n^{(4)}(t) \\ \tilde{\Xi}_n^{(5)}(t) \\ \tilde{\Xi}_n^{(6)}(t) \end{pmatrix} := \begin{pmatrix} \sqrt{n}\mathbb{P}_n^*(\epsilon \mathbf{1}_{Z \leq \zeta_0}) \\ \sqrt{n}\mathbb{P}_n^*(\epsilon_n \mathbf{1}_{Z > \zeta_0}) \\ n\mathbb{P}_n^*(\mathbf{1}_{\zeta_0 + \frac{t}{n} < Z \leq \zeta_0}) \\ n\mathbb{P}_n^*(\epsilon \mathbf{1}_{\zeta_0 + \frac{t}{n} < Z \leq \zeta_0}) \\ n\mathbb{P}_n^*(\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{t}{n}}) \\ n\mathbb{P}_n^*(\epsilon \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{t}{n}}) \end{pmatrix} \tag{17}
$$

and just as in that case, we can also define the process $\tilde{E}_n^*$ by

$$
\begin{aligned}
\tilde{E}_n^*(h) \;\; := \;\; & 2h_1 \tilde{\Xi}_n^{(1)}(h_3) - h_1^2 \mathbb{P}_n^*(Z \leq \zeta_0) + 2h_2 \tilde{\Xi}_n^{(2)}(h_3) - h_2^2 \mathbb{P}_n^*(Z > \zeta_0) \\
& + 2(\beta_0 - \alpha_0)\tilde{\Xi}_n^{(4)}(h_3) - (\alpha_0 - \beta_0)^2 \tilde{\Xi}_n^{(3)}(h_3) \\
& + 2(\alpha_0 - \beta_0)\tilde{\Xi}_n^{(6)}(h_3) - (\alpha_0 - \beta_0)^2 \Xi_n^{(5)}(h_3).
\end{aligned}
$$

Then, it can be shown that $\tilde{E}_n - \tilde{E}_n^* \xrightarrow{\mathbf{P}} 0$ in $\mathcal{D}_K$ for any compact rectangle $K \subset \mathbb{R}^3$ and that the sequence $(\tilde{E}_n^*)_{n=1}^{\infty}$ is tight in $\mathcal{D}_K$.

In what follows we will describe the limiting distribution of $\tilde{E}_n^*$, namely $\tilde{E}^*$, and show that the (unconditional) asymptotic distribution of $\tilde{\Delta}_n^*$ is that of the smallest argmax of $\tilde{E}^*$. This result will help us show that the ECDF bootstrap is inconsistent.

We start by introducing some notation. Recall the definitions of the random elements $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\nu_1$, $\nu_2$, $\mathbf{u}$ and $\mathbf{v}$ as in the discussion preceding (11). Also let $\tau = (\tau_n)_{n=1}^{\infty}$ and $\kappa = (\kappa_n)_{n=1}^{\infty}$ two sequences of i.i.d. Poisson(1) random variables. Assume, in addition, that $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\nu_1$, $\nu_2$, $\mathbf{v}$, $\mathbf{u}$, $\tau$ and $\kappa$ are all mutually independent. Then, define the process $\tilde{\Xi} = (\tilde{\Xi}^{(1)}, \ldots, \tilde{\Xi}^{(6)})'$ as

$$
\tilde{\Xi}(t) := \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \sum_{0 < j \leq \nu_1(-t)} \kappa_j \mathbf{1}_{t<0} \\ \sum_{0 < j \leq \nu_1(-t)} \upsilon_j \kappa_j \mathbf{1}_{t<0} \\ \sum_{0 < j \leq \nu_2(t)} \tau_j \mathbf{1}_{t \geq 0} \\ \sum_{0 < j \leq \nu_2(t)} u_j \tau_j \mathbf{1}_{t \geq 0} \end{pmatrix} \tag{18}
$$

for $t \in \mathbb{R}$ and let $\tilde{E}^*$ be given by

$$
\begin{aligned}
\tilde{E}^*(h) \;\; = \;\; & 2h_1 \tilde{\Xi}^{(1)}(h_3) - h_1^2 \mathbb{P}(Z \leq \zeta_0) + 2h_2 \tilde{\Xi}^{(2)}(h_3) - h_2^2 \mathbb{P}(Z > \zeta_0) \\
& + 2(\beta_0 - \alpha_0)\tilde{\Xi}^{(4)}(h_3) - (\alpha_0 - \beta_0)^2 \tilde{\Xi}^{(3)}(h_3) \\
& + 2(\alpha_0 - \beta_0)\tilde{\Xi}^{(6)}(h_3) - (\alpha_0 - \beta_0)^2 \tilde{\Xi}^{(5)}(h_3) \tag{19}
\end{aligned}
$$

19

for $h = (h_1, h_2, h_3) \in \mathbb{R}^3$. Additionally define the $\mathcal{S}$–valued (pure jump) processes $\tilde{J}_n$, $\tilde{J}_n^*$ and $\tilde{J}^*$ as

$$\tilde{J}_n^*(t) = \tilde{J}_n(t) \quad := \quad n\mathbb{P}_n^*(\mathbf{1}_{\zeta_0 + \frac{t}{n} < Z \leq \zeta_0}) + n\mathbb{P}_n^*(\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{t}{n}}), \tag{20}$$

$$\tilde{J}^*(t) \quad := \quad \nu_1(-t)\mathbf{1}_{t<0} + \nu_2(t)\mathbf{1}_{t\geq 0}. \tag{21}$$

Lemma 4.6 (proved in Section A.2.11) now states the asymptotic distribution of $\tilde{E}_n$ and of $n(\zeta_n^* - \zeta_0)$.

**Lemma 4.6** *Consider the processes $\tilde{\Xi}_n$, $\tilde{E}_n$, $\tilde{J}_n$, $\tilde{\Xi}$, $\tilde{E}^*$ and $\tilde{J}^*$ as defined in (17), (16), (20), (18), (19) and (21), respectively. Then, unconditionally,*

*(i) $\tilde{\Xi}_n \rightsquigarrow \tilde{\Xi}$ in $\mathbb{R}^2 \times \mathcal{D}_I^4$ for any compact interval $I \subset \mathbb{R}$;*

*(ii) $(\tilde{E}_n, \tilde{J}_n) \rightsquigarrow (\tilde{E}^*, \tilde{J}^*)$ in $\mathcal{D}_K \times \mathcal{S}_I$ for any compact interval $I \subset \mathbb{R}$ and any compact rectangle $K = A \times B \times I \subset \mathbb{R}^3$;*

*(iii) $\tilde{\Delta}_n^* = sargmax_{h \in \mathbb{R}^3}\{\tilde{E}_n(h)\} \rightsquigarrow sargmax_{h \in \mathbb{R}^3}\{\tilde{E}^*(h)\}$.*

*As a consequence, if the ECDF bootstrap is consistent, the variance of $sargmax_{h \in \mathbb{R}^3}\{\tilde{E}^*(h)\}$ must be twice that of $sargmax_{h \in \mathbb{R}^3}\{E^*(h)\}$.*

As analytic expressions for the asymptotic variances of $n(\zeta_n^* - \zeta_0)$ and $n(\hat{\zeta}_n - \zeta_0)$ are not known, we use simulations to compute them. As an illustration, we take $\epsilon \sim N(0,1)$, $Z \sim N(0,1)$, $\alpha_0 = -1$, $\beta_0 = 1$ and $\zeta_0 = 0$ in (1). We approximate the limiting variances with the sample variances computed from 20,000 observations from each of the two asymptotic distributions. Our results are summarized in the following table, which immediately shows that the asymptotic variance of $n(\zeta_n^* - \zeta_0)$ is not twice that of $n(\hat{\zeta}_n - \zeta_0)$. Thus the ECDF bootstrap cannot be consistent.

| Random variable | Asymptotic Variance |
|:---:|:---:|
| $n(\hat{\zeta}_n - \zeta_0)$ | 7.620948 |
| $n(\zeta_n^* - \zeta_0)$ | 63.98377 |

## 4.2   Scheme 2 (Bootstrapping "residuals")

Another resampling procedure that arises naturally in a regression setup is bootstrapping "residuals". As with scheme 1, bootstrapping the "residuals" fixing the covariates is also *inconsistent*. Heuristically speaking, the resampling distribution fails to approximate the density of the predictor at the change-point $\zeta_0$ at rate-$n$, and this leads to the inconsistency.

We recall the notation of Section 2. There we described the basic elements of the traditional fixed-design bootstrap of residuals and how to compute the bootstrap estimates $\theta_n^*$. We first show that these bootstrap estimators converge conditionally in probability (almost surely) to the true value of the parameter. Then, we will provide a strong argument against the consistency of this bootstrap scheme. For notational convenience, we introduce the process $R_n$ given by

$$R_n(\theta) := -\frac{1}{n}\sum_{j=1}^{n}\left(Y_{n,j}^* - \alpha\mathbf{1}_{Z_j\leq\zeta} - \beta\mathbf{1}_{Z_j>\zeta}\right)^2 \quad \forall\,\theta\in\Theta.$$

We start by showing that the "centered" empirical distribution for the least squares residuals, $\mathbb{P}_n^\epsilon$, converges to the distribution of $\epsilon$ in total variation distance with probability one and its second moment is an almost surely consistent estimator of $\sigma^2$. This lemma will also be useful for the analysis of the smoothed bootstrap procedure. The proof can be found in Section A.2.12.

**Lemma 4.7** *Let $G$ and $\varphi$ be, respectively, the distribution and characteristic functions of $\epsilon$. Then,*

*(i) for any $\eta > 0$ we have that $\displaystyle\sup_{|\xi|\leq\eta}\left\{\left|\int e^{i\xi x}d\mathbb{P}_n^\epsilon(x) - \varphi\left(\xi\right)\right|\right\} \xrightarrow{a.s.} 0$;*

*(ii) $\|\mathbb{P}_n^\epsilon - G\|_{\mathbb{R}} \xrightarrow{a.s.} 0$;*

*(iii) $\displaystyle\int x^2 d\mathbb{P}_n^\epsilon(x) \xrightarrow{a.s.} \sigma^2$;*

*(iv) $\displaystyle\int |x|d\mathbb{P}_n^\epsilon(x) \xrightarrow{a.s.} \mathbb{P}(|\epsilon|)$;*

*(v) if $\epsilon$ has a finite third moment under $\mathbb{P}$, then*

$$\varlimsup_{n\to\infty}\int|x|^3d\mathbb{P}_n^\epsilon(x) < \infty \quad \text{almost surely.}$$

The next result (proved in Section A.2.13) shows that the bootstrapped least squares estimators converge conditionally in probability with probability one.

**Proposition 4.2** *Let $K \subset \Theta$ be a compact rectangle. Then,*

*(i) $\|R_n + \mathbb{P}_n^*(\tilde{\epsilon}_n^2) - M_n - \sigma^2\|_K \xrightarrow[a.s.]{\mathbf{P}_{\mathfrak{X}}} 0$;*

*(ii) $\|R_n + \mathbb{P}_n^*(\tilde{\epsilon}_n^2) - M - \sigma^2\|_K \xrightarrow[a.s.]{\mathbf{P}_{\mathfrak{X}}} 0$;*

*(iii)* $\theta_n^* \xrightarrow[a.s.]{\mathbf{P_{\mathfrak{X}}}} \theta_0$ *and* $\theta_n^* - \hat{\theta}_n \xrightarrow[a.s.]{\mathbf{P_{\mathfrak{X}}}} 0$.

*where $M_n$ and $M$ are defined as in (3) and the subsequent paragraph.*

Consider the following process

$$\hat{E}_n(h) = -\sum_{j=1}^{n} \left( Y_{n,j}^* - \left( \hat{\alpha}_n + \frac{h_1}{\sqrt{n}} \right) \mathbf{1}_{Z_j \leq \hat{\zeta}_n + \frac{h_3}{n}} - \left( \hat{\beta}_n + \frac{h_2}{\sqrt{n}} \right) \mathbf{1}_{Z_j > \hat{\zeta}_n + \frac{h_3}{n}} \right)^2 + \sum_{j=1}^{n} (\epsilon_{n,j}^*)^2.$$

Then for $n$ large enough we have that

$$\left( \sqrt{n}(\alpha_n^* - \hat{\alpha}_n), \sqrt{n}(\beta_n^* - \hat{\beta}_n), n(\zeta_n^* - \hat{\zeta}_n) \right) = \operatorname*{sargmax}_{h \in \mathbb{R}^3} \left\{ \hat{E}_n(h) \right\}.$$

Next we argue that the sequence $(\hat{E}_n)_{n=1}^{\infty}$ does not have a weak limit in probability and therefore distributional convergence of their corresponding smallest minimizers seems unreasonable. We refer the reader to Section A.2.14 for a complete proof of the statement.

**Lemma 4.8** *There is a compact rectangle $K \subset \mathbb{R}^3$ such that the sequence of processes $(\hat{E}_n)_{n=1}^{\infty}$ does not have a weak limit in probability in $\mathcal{D}_K$.*

# 5 Consistent bootstrap procedures

Here we will prove that the "smoothed bootstrap" (scheme 3) and the $m$ out of $n$ bootstrap (scheme 4) procedures yield consistent methods for constructing confidence intervals around the parameters.

## 5.1 Scheme 3 (Smoothed Bootstrap)

To show that scheme 3 (smoothed bootstrap + bootstrapping residuals) achieves consistency we appeal to Propositions 3.1, 3.2 and 3.3 by proving that the regularity conditions (I)-(VIII) of Section 3 hold for this scheme. Recall the description of this bootstrap procedure given in Section 2. Let $\hat{f}_n$ and $\hat{F}_n$ be the estimated smoothed density and distribution function of $Z$, respectively. For $I := [c, d] \subset \mathbb{R}$, a compact interval such that $\zeta_0 \in (c, d)$, we require the following two properties of $\hat{f}_n$ and $\hat{F}_n$:

$$\|\hat{F}_n - F\|_{\mathbb{R}} \xrightarrow{a.s.} 0; \tag{22}$$

$$\|\hat{f}_n - f\|_I \xrightarrow{a.s.} 0. \tag{23}$$

We would want to highlight that these conditions are fulfilled by many density estimation procedures. In particular, they hold when the density $f$ is continuous

and we let $\hat{f}_n$ be the kernel density estimator constructed from a suitable choice of kernel and bandwidth (e.g., see Silverman (1978)).

Let $\theta_n = \hat{\theta}_n$, $m_n = n$ and $\mathbb{Q}_n$ be the distribution that generates the bootstrap sample. Observe that under $\mathbb{Q}_n$, $\tilde{\epsilon}_n$ and $Z$ are independent and that $Z$ is a continuous random variable with density $\hat{f}_n$. The next result (proved in Section A.2.15) shows that the bootstrapped least squares estimators achieve the right rate of convergence.

**Proposition 5.1** *If (22) and (23) hold, then w.p.1, the sequence of conditional distributions of $\left( \sqrt{n}(\alpha_n^* - \hat{\alpha}_n), \sqrt{n}(\beta_n^* - \hat{\beta}_n), n(\zeta_n^* - \hat{\zeta}_n) \right)$ is tight.*

Scheme 3 uses an approximation to the density of $Z$ and this turns out to be crucial. The bootstrap measures now satisfy property (VI) on Section 3 and the bootstrap procedure is *strongly consistent*, as shown in the next result (proved in Section A.2.16).

**Proposition 5.2** *For scheme 3, provided that (22) and (23) hold, conditions (I)–(VIII) are satisfied with probability one, and thus,*

$$
\begin{pmatrix} \sqrt{n}(\alpha_n^* - \hat{\alpha}_n) \\ \sqrt{n}(\beta_n^* - \hat{\beta}_n) \\ n(\zeta_n^* - \hat{\zeta}_n) \end{pmatrix} \rightsquigarrow \operatorname*{sargmax}_{h \in \mathbb{R}^3} \{E^*(h)\} \quad \text{almost surely.}
$$

## 5.2   Scheme 4 ($m$ out of $n$ bootstrap)

For this scheme we will again use the framework established in Section 3. We take $(m_n)_{n=1}^\infty$ to be any sequence of natural numbers which increases to infinity, $\hat{\theta}_n = \theta_n$ and $\mathbb{Q}_n = \mathbb{P}_n$. The next result (proved in Section A.2.17) shows the *weak consistency* of this procedure.

**Proposition 5.3** *If $m_n = o(n)$ and $m_n \to \infty$, then conditions (I)–(VIII) hold (in probability) and we have*

$$
\begin{pmatrix} \sqrt{m_n}(\alpha_n^* - \hat{\alpha}_n) \\ \sqrt{m_n}(\beta_n^* - \hat{\beta}_n) \\ m_n(\zeta_n^* - \hat{\zeta}_n) \end{pmatrix} \rightsquigarrow \operatorname*{sargmax}_{h \in \mathbb{R}^3} \{E^*(h)\} \quad \text{in probability.} \tag{24}
$$

**Remark:** To prove Proposition 5.3, we will, in fact, show that for every subsequence $(n_k)_{k=1}^\infty$, there is a further subsequence $(n_{k_s})_{s=1}^\infty$, such that (I)-(VIII) hold w.p. 1 for $(n_{k_s})_{s=1}^\infty$ and (24) holds almost surely along the subsequence $(n_{k_s})_{s=1}^\infty$.

# 6 Simulation experiments

In this section we report the finite sample performance of the different bootstrap schemes on simulated data. We simulated random draws from four different models following (1). Each of these corresponded to choosing different pairs $(F, G)$ of distributions for $Z$ and $\epsilon$ (having mean 0), respectively. The pairs considered were $(N(0, 2), N(0, 1))$, $(4B(4, 6) - 2, N(0, 1))$, $(4B(4, 6) - 2, \text{Unif}(-1, 1))$, and $(4B(4, 6) - 2, \Gamma(4, 2) - 2)$, where $B(\cdot, \cdot)$ and $\Gamma(\cdot, \cdot)$ denote the beta and gamma distributions respectively.

For each of these models, we considered 1000 random samples of sizes $n = 50, 100, 200, 500$. For each sample, and for each of the bootstrap schemes, we took $4n$ bootstrap replicates to approximate the bootstrap distribution. The following table provides the estimated coverage proportions of nominal 95% CIs and average lengths of the CIs obtained using the 4 different bootstrap schemes for each of the four models.

At this point, we want to make some remarks about the computation of the estimators. We used a kernel density estimator based on the Gaussian kernel and chose the bandwidth by the so-called "normal reference rule" (see Scott (1992), page 131). In the case of the $m$ out of $n$ bootstrap, we did not use any data driven choice of $m_n$, but tried 3 different possibilities: $\lceil n^{\frac{4}{5}} \rceil$, $\lceil n^{\frac{9}{10}} \rceil$ and $\lceil n^{\frac{14}{15}} \rceil$. We will refer to the fixed-design bootstrapping of residuals scheme by FDR.

| $Z \sim N(0, 2), \epsilon \sim N(0, 1)$ | | | | | | |
|---|---|---|---|---|---|---|
| Scheme | $n = 50$ | | $n = 200$ | | $n = 500$ | |
| | Coverage | Avg Length | Coverage | Avg Length | Coverage | Avg Length |
| ECDF | 0.83 | 1.14 | 0.79 | 0.22 | 0.81 | 0.08 |
| Smoothed | 0.94 | 0.94 | 0.95 | 0.19 | 0.95 | 0.07 |
| FDR | 0.83 | 0.76 | 0.86 | 0.16 | 0.90 | 0.06 |
| $\lceil n^{4/5} \rceil$ | 0.87 | 0.87 | 0.91 | 0.23 | 0.91 | 0.08 |
| $\lceil n^{9/10} \rceil$ | 0.85 | 1.02 | 0.87 | 0.21 | 0.87 | 0.079 |
| $\lceil n^{14/15} \rceil$ | 0.85 | 1.05 | 0.84 | 0.21 | 0.86 | 0.08 |
| $Z \sim 4B(4, 6) - 2, \epsilon \sim N(0, 1)$ | | | | | | |
| Scheme | $n = 50$ | | $n = 200$ | | $n = 500$ | |
| | Coverage | Avg Length | Coverage | Avg Length | Coverage | Avg Length |
| ECDF | 0.80 | 0.54 | 0.80 | 0.11 | 0.81 | 0.04 |
| Smoothed | 0.96 | 0.46 | 0.94 | 0.11 | 0.95 | 0.47 |
| FDR | 0.73 | 0.32 | 0.77 | 0.08 | 0.79 | 0.03 |
| $\lceil n^{4/5} \rceil$ | 0.88 | 0.53 | 0.89 | 0.11 | 0.90 | 0.04 |
| $\lceil n^{9/10} \rceil$ | 0.85 | 0.54 | 0.86 | 0.11 | 0.88 | 0.04 |
| $\lceil n^{14/15} \rceil$ | 0.83 | 0.55 | 0.84 | 0.11 | 0.87 | 0.04 |
| $Z \sim 4B(4, 6) - 2, \epsilon \sim \text{Unif}(-1, 1)$ | | | | | | |
| Scheme | $n = 50$ | | $n = 200$ | | $n = 500$ | |
| | Coverage | Avg Length | Coverage | Avg Length | Coverage | Avg Length |
| ECDF | 0.80 | 0.40 | 0.80 | 0.08 | 0.81 | 0.03 |
| Smoothed | 0.94 | 0.33 | 0.95 | 0.08 | 0.96 | 0.04 |
| FDR | 0.75 | 0.26 | 0.77 | 0.06 | 0.81 | 0.02 |
| $\lceil n^{4/5} \rceil$ | 0.88 | 0.36 | 0.88 | 0.09 | 0.91 | 0.04 |
| $\lceil n^{9/10} \rceil$ | 0.85 | 0.39 | 0.85 | 0.08 | 0.87 | 0.03 |
| $\lceil n^{14/15} \rceil$ | 0.83 | 0.39 | 0.84 | 0.08 | 0.85 | 0.03 |

| $Z \sim 4B(4,6) - 2, \epsilon \sim \Gamma(4,2) - 2$ | | | | | | |
|---|---|---|---|---|---|---|
| Scheme | $n = 50$ | | $n = 200$ | | $n = 500$ | |
| | Coverage | Avg Length | Coverage | Avg Length | Coverage | Avg Length |
| ECDF | 0.80 | 0.49 | 0.80 | 0.09 | 0.81 | 0.04 |
| Smoothed | 0.93 | 0.36 | 0.95 | 0.08 | 0.96 | 0.03 |
| FDR | 0.76 | 0.30 | 0.77 | 0.06 | 0.80 | 0.02 |
| $\lceil n^{4/5} \rceil$ | 0.87 | 0.43 | 0.88 | 0.10 | 0.91 | 0.03 |
| $\lceil n^{9/10} \rceil$ | 0.85 | 0.46 | 0.84 | 0.09 | 0.88 | 0.03 |
| $\lceil n^{14/15} \rceil$ | 0.83 | 0.48 | 0.85 | 0.09 | 0.85 | 0.03 |

We can see from the table that the smoothed bootstrap scheme outperforms all the others in terms of coverage. It must also be noted that this is achieved without a relative increase in the lengths of the intervals. The $m$ out of $n$ bootstrap with $\lceil n^{4/5} \rceil$ also performs reasonably well. It clearly outperforms all other $m$ out of $n$ schemes as well as ECDF and FDR bootstrap procedures (which are inconsistent). Figure 1 shows the histograms of the distribution of $n(\hat{\zeta}_n - \zeta_0)$ (obtained from 1000 random samples) and its bootstrap estimates obtained from the 4 different bootstrap schemes (using 2000 bootstrap samples each) from a single data set of size $n = 500$ from model (1) with $Z \sim 4B(4,6) - 2, \epsilon \sim \Gamma(4,2) - 2, \alpha_0 = -1, \beta_0 = 1, \zeta_0 = 0$. The histograms clearly show that the smoothed bootstrap (top right panel) provides, by far, the best approximation to both, the actual (top middle panel) and the limiting distributions (top left panel). In fact, the histograms of the distribution of $n(\hat{\zeta}_n - \zeta_0)$ and the corresponding smoothed bootstrap estimate are almost indistinguishable. The $m$ out of $n$ approach, although guaranteed to converge, lacks the efficiency of the smoothed bootstrap. This may be due to the fact that we do not have an optimal way of choosing the tuning parameter $m_n$. The smoothed bootstrap also requires the choice of a tuning parameter, namely, the smoothing bandwidth, but the in our analysis the results were very insensitive to the choice of the bandwidth. This is certainly an advantage for the smoothed bootstrap procedure.

# 7 More general change-point regression models

In this section we mention some of the broader implications of our analysis of (1) in the context of more general change-point models in regression. We can consider a model of the form

$$Y = \psi_{\alpha_0}(W, Z)\mathbf{1}_{Z \leq \zeta_0} + \xi_{\beta_0}(W, Z)\mathbf{1}_{Z > \zeta_0} + \epsilon, \tag{25}$$

where $Z$ is a continuous random variable; $W$ is a random vector of covariates; $\alpha_0 \in \mathbb{R}^p$ and $\beta_0 \in \mathbb{R}^q$ are two unknown Euclidian parameters; $\psi_\alpha(w, z)$ and $\xi_\beta(w, z)$ are known real-valued functions continuous in $(w, z)$ and twice continuously differ-
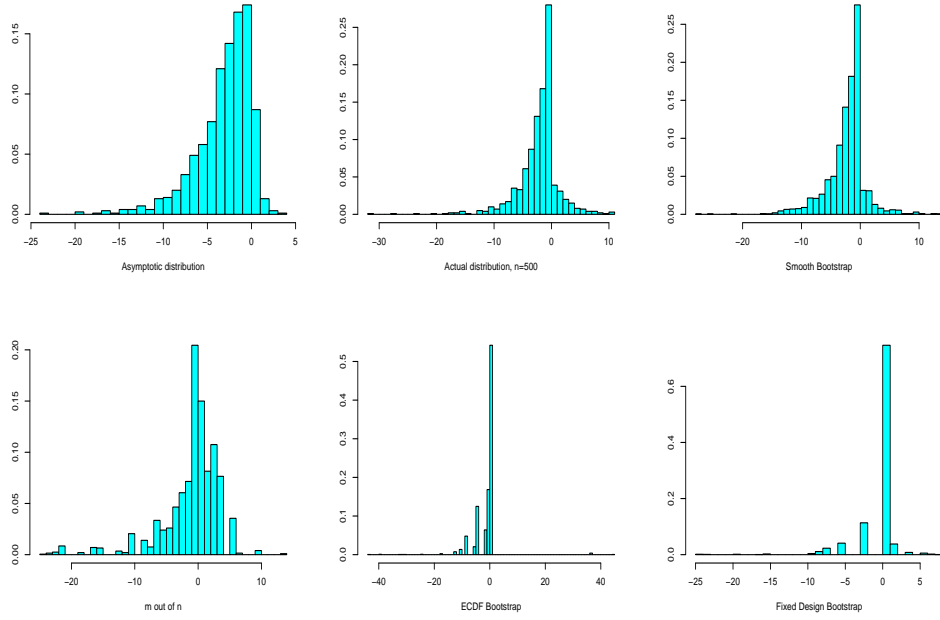
Figure 1: Histograms of the distribution of $n(\hat{\zeta}_n - \zeta_0)$ and its bootstrap estimates: the asymptotic distribution of $n(\hat{\zeta}_n - \zeta_0)$ (top left); the actual distribution of $n(\hat{\zeta}_n - \zeta_0)$ (top middle); the distribution of $n(\zeta_n^* - \hat{\zeta}_n)$ for the smoothed (top right), ECDF (bottom middle) and FDR (bottom right) schemes; the distribution of $m_n(\zeta_n^* - \hat{\zeta}_n)$, $m_n = \lceil n^{\frac{4}{5}} \rceil$ (bottom left).

entiable in $\alpha$ and $\beta$ respectively; $\zeta_0 \in [a,b] \subset \mathrm{supp}(Z) \subset \mathbb{R}$ is the change-point; $\epsilon$ is a continuous random variable, independent of $(W,Z)$ with zero expectation and finite variance $\sigma^2 > 0$. We assume that $\psi_{\alpha_0}(W,Z)$ is identifiable from $\xi_{\beta_0}(W,Z)$ and that the least squares problems

$$\min_{\alpha \in \mathbb{R}^p} \left\{ \sum_{Z_j \leq \zeta} (Y_j - \psi_\alpha(W_j, Z_j))^2 \right\} \quad \text{and} \quad \min_{\beta \in \mathbb{R}^q} \left\{ \sum_{Z_j > \zeta} (Y_j - \xi_\beta(W_j, Z_j))^2 \right\}$$

are well-posed for every possible data set $\{(Y_1, Z_1, W_1), \ldots, (Y_n, Z_n, W_n)\}$ and any $\zeta \in \mathrm{supp}(Z)^\circ$. We also assume that $\psi_{\alpha_0}(w, \zeta_0) \neq \xi_{\beta_0}(w, \zeta_0)$ for every value of $w$. Like in the simple case, the method of least squares can be used to compute estimators $\hat{\alpha}_n$, $\hat{\beta}_n$ and $\hat{\zeta}_n$. One simply takes the minimizer $(\hat{\alpha}_n, \hat{\beta}_n, \hat{\zeta}_n)$ of

$$\sum_{j=1}^n \left( Y_j - \psi_\alpha(W_j, Z_j) \mathbf{1}_{Z_j \leq \zeta} + \xi_\beta(W_j, Z_j) \mathbf{1}_{Z_j > \zeta} \right)^2$$

with the smallest $\zeta$-component.

26

Since the simple model (1) is a particular case of (25), one can immediately conclude from our analysis that the usual ECDF and residual bootstrap procedures will not be consistent. However, the smoothed bootstrap can be adapted to produce consistent interval estimation. The modified scheme can be described as follows:

1. Choose some procedure (e.g., kernel density estimation) to build a distribution $\hat{F}_n$ with density $\hat{f}_n$ such that $\hat{f}_n \to f$ uniformly on some open interval containing $\zeta_0$ w.p. 1, where $f$ is the density of $Z$. Let $\mathbb{P}_n^\epsilon$ and $\mathbb{P}_n^W$ be the empirical measures of the centered residuals (as in the description of Scheme 2 in Section 2) and $W_1, \ldots, W_n$, respectively.

2. Get i.i.d. replicates $Z_{n,1}^*, \ldots, Z_{n,n}^*$ from $\hat{F}_n$ and sample, independently, $\epsilon_{n,1}^*, \ldots, \epsilon_{n,n}^* \overset{i.i.d.}{\sim} \mathbb{P}_n^\epsilon$ and $W_{n,1}^*, \ldots, W_{n,n}^* \overset{i.i.d.}{\sim} \mathbb{P}_n^W$. We could have also kept the $W_i$'s fixed, i.e., $W_{n,i}^* = W_i$.

3. Define $Y_{n,j}^* = \psi_{\hat{\alpha}_n}(W_{n,j}^*, Z_{n,j}^*)\mathbf{1}_{Z_{n,j}^* \le \hat{\zeta}_n} + \xi_{\hat{\beta}_n}(W_{n,j}^*, Z_{n,j}^*)\mathbf{1}_{Z_{n,j}^* > \hat{\zeta}_n} + \epsilon_{n,j}^*$ for all $j = 1, \ldots, n$.

4. Compute the bootstrap least squares estimators $(\alpha_n^*, \beta_n^*, \zeta_n^*)$ by taking the minimizer of

$$\sum_{j=1}^n \left(Y_{n,j}^* - \psi_\alpha(W_{n,j}^*, Z_{n,j}^*)\mathbf{1}_{Z_{n,j}^* \le \zeta} - \xi_\beta(W_{n,j}^*, Z_{n,j}^*)\mathbf{1}_{Z_{n,j}^* > \zeta}\right)^2$$

with the smallest $\zeta$-component.

5. Approximate the distribution of $n(\hat{\zeta}_n - \zeta_0)$ with the (conditional) distribution of $n(\zeta_n^* - \hat{\zeta}_n)$.

Although our analysis indicates that this smoothed bootstrap procedure must be consistent, it is difficult to use our methods to prove consistency in such generality. However, the proof of consistency for the simple model (1) can be adapted to cover the case of parametric additive models, i.e., when $\psi_\alpha(w, z)$ and $\xi_\beta(w, z)$ are of the form

$$\psi_\alpha(w, z) = \sum_{j=1}^p \alpha_j g_j(w, z), \quad \text{and} \quad \xi_\beta(w, z) = \sum_{k=1}^q \beta_k h_k(w, z),$$

where $g_j, h_k$, $j = 1, \ldots, p$, $k = 1, \ldots, q$ are known smooth functions.

# 8  Acknowledgements

We would like to thank Souvik Ghosh for his helpful comments on the proof of Lemma 4.6.

# A  Appendix

In this appendix we provide the proofs of most of the results stated in the previous sections. We start with some results that characterize convergence in the space $\mathcal{D}_K$ with metric $d_K$.

## A.1  The space $\mathcal{D}_K$

Recall that $\mathcal{D}_K$ is the space of all functions on $K \subset \mathbb{R}^3$ that are continuous in the first two co-ordinates and càdlàg in the third. We keep the notation introduced in Section 3.

**Lemma A.1** *Let $K = A \times I \subset \mathbb{R}^3$ be a compact rectangle where $A \subset \mathbb{R}^2$ and $I \subset \mathbb{R}$. Let $W$ a continuous function on $K$ and $(W_n)_{n=1}^\infty$ a sequence of elements in $\mathcal{D}_K$ such that $d_K(W_n, W) \to 0$. Then, with the notation $\|x\|_K = \sup\{|x(h)| : h \in K\}$, we have that $\|W_n - W\|_K \to 0$.*

**Proof:**  Let $\epsilon > 0$. Since $K$ is compact, $W$ is uniformly continuous on $K$ and therefore there is $\delta > 0$ such that $|W(\theta) - W(\vartheta)| < \frac{\epsilon}{2}$ whenever $|\theta - \vartheta| < \delta$. Also, there is $\rho > 0$ such that $\sup\{|s - \lambda(s)| : s \in I\} < \delta$ whenever $\|\lambda\| < \rho$. It suffices to choose $\rho < \frac{1}{4} \wedge \frac{\delta}{2L}$ where $L$ is the length of $I$. To see this, assume $I = [a, b]$ and observe that for any $\tau \in (0, \frac{1}{4})$, $\tau < 2\tau - 4\tau^2 \leq \log(1 + 2\tau)$ and for any $\tau > -1$, $\log(1 + \tau) \leq \tau$. It follows that for any $s \in I$, $\log(1 - 2\rho) < -\rho \leq \log\frac{\lambda s - a}{s - a} \leq \rho < 2\rho - 4\rho^2 \leq \log(1 + 2\rho)$ and thus, $|\lambda s - s| < 2(s - a)\rho \leq 2L\rho$.

Now, since $d_K(W_n, W) \to 0$, there is $N \in \mathbb{N}$ such that for any $n \geq N$ there exists $\lambda_n \in \Lambda_I$ with the property that $\|\lambda_n\| < \rho \wedge \frac{\epsilon}{2}$ and

$$\sup_{(\alpha, \beta, \zeta) \in K} \{|W_n(\alpha, \beta, \zeta) - W(\alpha, \beta, \lambda_n \zeta)|\} < \rho \wedge \frac{\epsilon}{2}. \tag{26}$$

Then, for any $\theta = (\alpha, \beta, \zeta) \in K$ and any $n \geq N$ we have that $|(\alpha, \beta, \lambda_n(\zeta)) - (\alpha, \beta, \zeta)| = |\lambda_n(\zeta) - \zeta| < \delta$ and thus, we can bound $|W_n(\theta) - W(\theta)|$, by

$$|W_n(\alpha, \beta, \zeta) - W(\alpha, \beta, \lambda_n(\zeta))| + |W(\alpha, \beta, \zeta) - W(\alpha, \beta, \lambda_n(\zeta))|$$
$$< \epsilon/2 + \epsilon/2 = \epsilon$$

using (26) and the uniform continuity of $W$. From this it follows that $\|W_n - W\| \leq \epsilon$ for any $n \geq N$. $\qquad\square$

Lemma A.1 shows that as long as the limit is continuous, convergence in the uniform and Skorohod topologies are equivalent. The next result concerns the continuity of the smallest argmax functional.

**Lemma A.2** *Let $K \subset \mathbb{R}^3$ be a compact rectangle and $W \in \mathcal{D}_K$ be a continuous function which has a unique maximizer $\theta^* \in K$. Then, the smallest argmax functional is continuous at $W$ (with respect to both, the uniform and Skorohod topologies).*

**Proof:** Let $(W_n)_{n=1}^\infty$ be a sequence converging to $W$ in the Skorohod topology. Let $\epsilon > 0$ be given. Let $G$ be the open ball of radius $\epsilon$ around $\theta^*$ and let $\delta := \left(W(\theta^*) - \sup_{\theta \in K \backslash G}\{W(\theta)\}\right)/2 > 0$. By Lemma A.1 we have $\|W_n - W\|_K < \delta$ for all large $n$. If this condition is satisfied, then

$$W(\theta^*) = 2\delta + \sup_{\theta \in K \backslash G}\{W(\theta)\} > \delta + \sup_{\theta \in K \backslash G}\{W_n(\theta)\}.$$

But $\|W_n - W\|_K < \delta$ also implies that $\sup_{\theta \in K}\{W_n(\theta)\} > W(\theta^*) - \delta$. The combination of these two facts shows that if $\|W_n - W\|_K < \delta$, then any maximizer of $W_n$ must belong to $G$. Thus, $|\text{sargmax}_{\vartheta \in K}\{W_n(\vartheta)\} - \theta^*| < \epsilon$ for $n$ large enough. $\square$

### A.1.1 A convergence theorem for the smallest argmax functional

Recall the definitions of $\mathcal{S}$ and $\mathcal{S}_I$, where $I \subset \mathbb{R}$ is any interval containing 0, which were provided in Section 3. For a compact rectangle $K = I_1 \times I_2 \times I_3 \subset \mathbb{R}^3$ containing the origin, consider the subspace $\mathcal{D}_K^0$ of $\mathcal{D}_K$ consisting of all functions $\psi \in \mathcal{D}_K$ which can be expressed as:

$$\psi(h_1, h_2, h_3) = V_0(h_1, h_2)\mathbf{1}_{a_{-1} \leq h_3 < a_1} + \sum_{k=1}^\infty V_k(h_1, h_2)\mathbf{1}_{a_k \leq h_3 < a_{k+1}}$$

$$+ \sum_{k=1}^\infty V_{-k}(h_1, h_2)\mathbf{1}_{a_{-k-1} \leq h_3 < a_{-k}} \qquad (27)$$

where $(\ldots < a_{-k-1} < a_{-k} < \ldots < a_0 = 0 < \ldots < a_k < a_{k+1} < \ldots)_{k \in \mathbb{N}}$ is a sequence of jumps and $(V_k)_{k \in \mathbb{Z}}$ is a collection of continuous functions. We write $\mathcal{D}^0$ when $K = \mathbb{R}^3$. Observe that the representation in (27) is not unique. However, knowledge

29

of the function $\psi$ and of the jumps $(a_k)$ completely determines the continuous functions $(V_k)_{k\in\mathbb{Z}}$. Associate with every $\psi$, expressed as in (27), a pure jump function $\tilde{\psi} \in \mathcal{S}$ whose sequence of jumps is exactly the $a_k$'s, i.e.,

$$\tilde{\psi}(t) = \sum_{k=1}^{\infty} \mathbf{1}_{a_k \le t} + \sum_{k=1}^{\infty} \mathbf{1}_{a_{-k} > t}. \tag{28}$$

Finally, we denote by sargmax and largmax the smallest and largest argmax functionals, respectively.

The next lemma, which mimics Lemma 3.1 of Lan et al. (2009), makes a statement about the continuity of the smallest argmax functional on the space $\mathcal{D}_K^0 \times \mathcal{S}_{I_3}$.

**Lemma A.3** *Let $C \in \mathbb{N}$, $K = [-C, C]^3$ and $\left(\psi_n, \tilde{\psi}_n\right)_{n=1}^{\infty}$, $(\psi_0, \tilde{\psi}_0)$ be functions in $\mathcal{D}_K^0 \times \mathcal{S}_{[-C,C]}$ such that $\psi_n$ satisfies (27) for the sequence of jumps of $\tilde{\psi}_n$ for any $n \ge 0$. Assume that $(\psi_n, \tilde{\psi}_n) \to (\psi_0, \tilde{\psi}_0)$ in $\mathcal{D}_K^0 \times \mathcal{S}_{[-C,C]}$ (with the product topology). Suppose, in addition, that $\psi_0$ can be expressed as (27) for the sequence of jumps $(\ldots < a_{-k-1} < a_{-k} < \ldots < a_0 = 0 < \ldots < a_k < a_{k+1} < \ldots)_{k\in\mathbb{N}}$ of $\tilde{\psi}_0$ and some strictly concave functions $(V_j)_{j\in\mathbb{Z}}$ with the property that for any finite subset $A \subset \mathbb{Z}$ there is only one $j \in A$ for which*

$$\max_{m \in A} \left\{ \sup_{h \in K} \{V_m(h_1, h_2)\} \right\} = \sup_{h \in K} \{V_j(h_1, h_2)\}. \tag{29}$$

*Then, $\underset{h \in K}{sargmax}\{\psi_n(h)\}$ and $\underset{h \in K}{largmax}\{\psi_n(h)\}$ are well-defined for sufficiently large $n \in \mathbb{N}$ and*

*(i) $\underset{h \in K}{sargmax}\{\psi_n(h)\} \to \underset{h \in K}{sargmax}\{\psi_0(h)\}$ as $n \to \infty$*

*(ii) $\underset{h \in K}{largmax}\{\psi_n(h)\} \to \underset{h \in K}{largmax}\{\psi_0(h)\}$ as $n \to \infty$.*

**Proof:** We can write $\psi_n$ in the form (27) with $(\ldots < a_{n,-k-1} < a_{n,-k} < \ldots < a_{n,0} = 0 < \ldots < a_{n,k} < a_{n,k+1} < \ldots)_{k\in\mathbb{N}}$ being the sequence of jumps of $\psi_n$ and $V_{n,j}$ being the continuous functions. Consequently, $\tilde{\psi}_n$, the pure jump process associated with $\psi_n$, can be expressed as (28) with jumps at $(a_{n,k})_{k\in\mathbb{Z}}$.

Let $N_r$ and $N_l$ be the number of jumps of $\tilde{\psi}_0$ in $[0, C]$ and $[-C, 0)$ respectively. Let $\rho > 0$ be sufficiently small such that all the points of the form $a_j \pm \rho$ are continuity points of $\psi_0$, for $-N_l \le j \le N_r$. Since convergence in the Skorohod topology of $\tilde{\psi}_n$ to $\tilde{\psi}_0$ implies point-wise convergence for continuity points of $\tilde{\psi}_0$ (see page 121 of Billingsley (1968)), and all of them are integer-valued functions, we see that

$\tilde{\psi}_n(a_j - \rho) = j - 1$ and $\tilde{\psi}_n(a_j + \rho) = j$ for any $1 \leq j \leq N_r$, and $\tilde{\psi}_n(C) = N_r$ for all sufficiently large $n$. Thus, for all but finitely many $n$'s we have that $\tilde{\psi}_n$ has exactly $N_r$ jumps between $0$ and $C$ and that the location of the $j$-th jump to the right of $0$ satisfies $|a_{n,j} - a_j| < \rho$. The same happens to the left of zero: for all but finitely many $n$'s, $\tilde{\psi}_n$ has exactly $N_l$ jumps in $[-C, 0)$ and the sequences of jumps $(a_{n,-j})_{n=1}^\infty$, $1 \leq j \leq N_l$, converge to the corresponding jumps $a_{-j}$.

Let $V^* = \sup \{V_j(h_1, h_2) : h \in K, -N_l \leq j \leq N_r\}$. Since all the $V_j$'s are continuous and $K$ is compact, this supremum is actually achieved at some value $(h_1^*, h_2^*) \in [-C, C]^2$. By (29) and the strict concavity of the $V_j$'s, it is seen that $(h_1^*, h_2^*)$ is unique and that there is a unique "flat stretch" at which this supremum is attained. Suppose, without loss of generality, that the maximum value is achieved in an interval of the form $[a_k, a_{k+1} \wedge C)$ for a unique $k \in \{1, \ldots, N_r\}$. Now, write $b_0 = 0$; $b_j = \frac{a_j + C \wedge a_{j+1}}{2}$ for $1 \leq j \leq N_r$; and $b_j = \frac{a_j + (-C) \vee a_{j-1}}{2}$ for $-N_l \leq j \leq -1$. Note that the $b_j$'s (for any value of the first two variables) are continuity points of both $\psi_0$ and $\tilde{\psi}_0$.

Let $\kappa = \min_{-N_l \leq j \leq N_r+1}(C \wedge a_j - (-C) \vee a_{j-1})$ be the length of the shortest stretch. Take $0 < \eta, \delta < \kappa/4$. Considering the convergence of the jumps of $\psi_n$ to those of $\psi_0$, there is $N \in \mathbb{N}$ such that for any $n \geq N$, the following two statements hold:

(a) Consider $\rho > 0$ such that if $\|\lambda\| < \rho$, then

$$\sup \{|s - \lambda(s)| : s \in [-C, C]\} < \delta,$$

just as in the proof of Lemma A.1. By the convergence of $\psi_n$ to $\psi_0$ in the Skorohod topology, there exists $\lambda_n \in \Lambda_{[-C,C]}$ such that $\|\lambda_n\| < \rho$ and

$$\sup_{h \in K} \{|\psi_n(h_1, h_2, \lambda_n(h_3)) - \psi_0(h_1, h_2, h_3)|\} < \eta.$$

(b) For any $1 \leq j \leq N_r$ (respectively $j = 0$, $-N_l \leq j \leq -1$), $b_j$ lies somewhere inside the interval $(a_{n,j} + \delta, C \wedge a_{n,j+1} - \delta)$ (respectively $(a_{n,-1} + \delta, \ a_{n,1} - \delta)$, $((-C) \vee a_{n,j-1} + \delta, a_{n,j} - \delta))$. This follows from what was proven in the first two paragraphs of this proof.

From (a) we see that $|\lambda_n(b_j) - b_j| < \delta$ for all $-N_l \leq j \leq N_r$. But (b) and the size of $\delta$ in turn imply that $b_j$ and $\lambda_n(b_j)$ belong to the same "flat stretch" of $\psi_n$ and thus $\psi_n(h_1, h_2, \lambda_n(b_j)) = \psi_n(h_1, h_2, b_j) = V_{n,j}(h_1, h_2)$ for all $(h_1, h_2) \in [-C, C]^2$ and all $-N_l \leq j \leq N_r$. Considering again (b) and the second inequality in (a), we conclude that $\|V_{n,j} - V_j\|_{[-C,C]^2} < \eta$ for all $-N_l \leq j \leq N_r$ and all $n \geq N$. Hence,

31

all the sequences $(V_{n,j})_{n=1}^{\infty}$ converge uniformly in $[-C,C]^2$ to their corresponding $V_j$. Consequently:

$$
\max_{\substack{-N_l \leq j \leq N_r \\ j \neq k}} \left\{ \sup_{h_1,h_2 \in [-C,C]} V_{n,j}(h_1,h_2) \right\} \longrightarrow \max_{\substack{-N_l \leq j \leq N_r \\ j \neq k}} \left\{ \sup_{h_1,h_2 \in [-C,C]} V_j(h_1,h_2) \right\},
$$

$$
\max_{h_1,h_2 \in [-C,C]} \{V_{n,k}(h_1,h_2)\} \longrightarrow \max_{h_1,h_2 \in [-C,C]} \{V_k(h_1,h_2)\} = V_k(h_1^*,h_2^*),
$$

$$
\operatorname*{argmax}_{h_1,h_2 \in [-C,C]} \{V_{n,k}(h_1,h_2)\} \longrightarrow \operatorname*{argmax}_{h_1,h_2 \in [-C,C]} \{V_k(h_1,h_2)\} = (h_1^*,h_2^*),
$$

$$
\overline{\lim} \max_{\substack{-N_l \leq j \leq N_r \\ j \neq k}} \left\{ \sup_{h_1,h_2 \in [-C,C]} V_{n,j}(h_1,h_2) \right\} < \underline{\lim} \max_{h_1,h_2 \in [-C,C]} \{V_{n,k}(h_1,h_2)\}.
$$

The above, together with (29) and the fact that $a_{n,k} \to a_k$ and $a_{n,k+1} \to a_{k+1}$, imply that

$$
\operatorname*{sargmax}_{h \in K} \{\psi_n(h)\} \to (h_1^*,h_2^*,a_k) = \operatorname*{sargmax}_{h \in K} \{\psi_0(h)\}
$$

$$
\operatorname*{largmax}_{h \in K} \{\psi_n(h)\} \to (h_1^*,h_2^*,a_{k+1}) = \operatorname*{largmax}_{h \in K} \{\psi_0(h)\}
$$

as $n \to \infty$. $\qquad\square$

## A.2   Some useful lemmas and proofs

We first give an account of a series of technical lemmas which will aid us in the proof of Propositions 3.1, 3.2 and 3.3.

**Lemma A.4** *Let $\alpha \neq \beta \in \mathbb{R}$. Consider the class of functions from $\mathbb{R}^2$ to $\mathbb{R}$ given by*

$$
\mathscr{A} = \left\{ \phi(y,z) := (y - \alpha \mathbf{1}_{(-\infty,\zeta]}(z) - \beta \mathbf{1}_{(\zeta,\infty]}(z))\mathbf{1}_I(z) | \zeta \in \mathbb{R}, I \subset \mathbb{R} \text{ is an interval} \right\}.
$$

*Then, $\mathscr{A}$ is a VC-subgraph class with envelope $|y| + |\alpha| + |\beta|$. There is an upper bound for the VC-index of $\mathscr{A}$ that is independent of $\alpha$ and $\beta$. Moreover, there is a continuous, increasing function $J_{\mathscr{A}}$ with $J_{\mathscr{A}}(1) < \infty$, which is also independent of $\alpha$ and $\beta$, and satisfies the following property: If $\mathscr{D} \subset \mathscr{A}$ is a subclass with envelope $B$ and $W_1, \ldots, W_n$ is a random sample, defined on some probability space $(\Omega, \mathcal{A}, \mathbf{P})$, from a distribution $\mu$ for which $\mu(B^2) < \infty$ and $\mu_n$ is the empirical measure defined by the sample, then*

$$
\int \sup_{\varphi \in \mathscr{D}} \{|(\mu_n - \mu)(\varphi)|\} \, d\mathbf{P} \leq \frac{J_{\mathscr{A}}(1)}{\sqrt{n}} \sqrt{\mu(B^2)}.
$$

**Proof:** We use the same notation as in Lemmas 2.6.17 and 2.6.18, page 147 of Van der Vaart and Wellner (1996). Consider the classes of functions $\mathscr{H} = \{y - \alpha\mathbf{1}_{(-\infty,\zeta]}(z) - \beta\mathbf{1}_{(\zeta,\infty]}(z) : \zeta \in \mathbb{R}\}$ and $\mathscr{K} = \{\mathbf{1}_{(-\infty,\zeta]}(z) : \zeta \in \mathbb{R}\}$. Then, $\mathscr{K}$ is a VC class with VC-index 2. It follows that $\mathscr{H} = (\beta - \alpha) \cdot \mathscr{K} + (y - \beta)$ is also VC. Recall that $\mathcal{F} = \{\mathbf{1}_I(z) : I \subset \mathbb{R} \text{ is an interval}\}$. Letting $[\varphi > t] := \{(y,z,t) : \varphi(y,z) > t\}$ for $\varphi \in \mathscr{A}$, we see that

$$\{[\varphi > t] : \varphi \in \mathscr{A}\} = \left(\mathbb{R} \times \{\mathcal{F} \leq 0\} \times (-\infty, 0)\right) \bigsqcup$$

$$\left(\{[\psi > t] : \psi \in \mathscr{H}\} \sqcap (\mathbb{R} \times \{\mathcal{F} > 0\} \times \mathbb{R})\right)$$

from which it follows that $\mathscr{A}$ is VC. Moreover, the VC-indexes of $\mathscr{K}$ and $\mathcal{F}$ are two and three for any choice of $\alpha$ and $\beta$. Hence, the corresponding VC-indexes of $\mathscr{H}$ and $\mathscr{A}$ both have upper bounds independent of $\alpha$ and $\beta$. The existence of the function $J_{\mathscr{A}}$ is a consequence of the maximal inequality 3.1 in Kim and Pollard (1990). Note that $J_{\mathscr{A}}$ only depends on the VC-index of the class $\mathscr{A}$, which in turn has an upper bound independent of $\alpha$ and $\beta$. $\qquad\square$

**Lemma A.5** *Suppose that (I)-(IV) hold. Then,*

*(i)* $\left\|\mathbb{Q}_n(\tilde{\epsilon}_n^2 \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \sigma^2 \mathbb{P}(Z \leq (\cdot) \wedge \zeta_0)\right\|_{[a,b]} \to 0,$

*(ii)* $\left\|\mathbb{Q}_n(|\tilde{\epsilon}_n| \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(|\epsilon|)\mathbb{P}(Z \leq (\cdot) \wedge \zeta_0)\right\|_{[a,b]} \to 0,$

*(iii)* $\left\|\mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n})\right\|_{[a,b]} \to 0,$ *and*

*(iv)* $\left\|\mathbb{Q}_n(\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\right\|_{[a,b]} \to 0.$

*Also, these statements are true if $\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}$ is replaced by any of $\mathbf{1}_{(\cdot) < Z \leq \zeta_n}$, $\mathbf{1}_{\zeta_n < Z \leq (\cdot)}$ or $\mathbf{1}_{Z > (\cdot) \vee \zeta_n}$.*

**Proof:** Since $\zeta_n \to \zeta_0$ and $Z$ is continuous, for any $\zeta \in [a,b]$, we obtain

$$
\begin{aligned}
\left|\mathbb{P}(Y^2 \mathbf{1}_{Z \leq \zeta \wedge \zeta_n}) - \mathbb{P}(Y^2 \mathbf{1}_{Z \leq \zeta \wedge \zeta_0})\right| &\leq \mathbb{P}(Y^2 |\mathbf{1}_{Z \leq \zeta_n} - \mathbf{1}_{Z \leq \zeta_0}|) \to 0, \\
\left|\mathbb{P}(|Y - \alpha_0| \mathbf{1}_{Z \leq \zeta \wedge \zeta_n}) - \mathbb{P}(|Y - \alpha_0| \mathbf{1}_{Z \leq \zeta \wedge \zeta_0})\right| &\leq \mathbb{P}(|Y| |\mathbf{1}_{Z \leq \zeta_n} - \mathbf{1}_{Z \leq \zeta_0}|) \to 0, \\
\left|\mathbb{P}(Y \mathbf{1}_{Z \leq \zeta \wedge \zeta_n}) - \mathbb{P}(Y \mathbf{1}_{Z \leq \zeta \wedge \zeta_0})\right| &\leq \mathbb{P}(|Y| |\mathbf{1}_{Z \leq \zeta_n} - \mathbf{1}_{Z \leq \zeta_0}|) \to 0, \\
\left|\mathbb{P}(\mathbf{1}_{Z \leq \zeta \wedge \zeta_n}) - \mathbb{P}(\mathbf{1}_{Z \leq \zeta \wedge \zeta_0})\right| &\leq \mathbb{P}(|\mathbf{1}_{Z \leq \zeta_n} - \mathbf{1}_{Z \leq \zeta_0}|) \to 0.
\end{aligned}
$$

Also note that the convergence is uniform in $\zeta \in [a, b]$. Thus,

$$\left\|\mathbb{Q}_n(Y^2 \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(Y^2 \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\right\|_{[a,b]} \leq \|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{H}}$$
$$+ \left\|\mathbb{P}(Y^2 \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(Y^2 \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\right\|_{[a,b]} \to 0$$

as $n \to \infty$ by (III). Similarly, we also obtain that $\|\mathbb{Q}_n(|Y - \alpha_0| \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(|Y - \alpha_0| \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\|_{[a,b]} \to 0$, $\|\mathbb{Q}_n(Y \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(Y \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\|_{[a,b]} \to 0$ and $\|\mathbb{Q}_n(\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\|_{[a,b]} \to 0$. This proves $(iv)$.

Finally, $(i)$, $(ii)$ and $(iii)$ follow as consequence of the convergence $\alpha_n \to \alpha_0$ and of the following inequalities:

$$\left\|\mathbb{Q}_n(\tilde{\epsilon}_n^2 \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \sigma^2 \mathbb{P}(Z \leq (\cdot) \wedge \zeta_0)\right\|_{[a,b]}$$
$$\leq \left\|\mathbb{Q}_n(Y^2 \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(Y^2 \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\right\|_{[a,b]} + 2|\alpha_n - \alpha_0|\mathbb{Q}_n(|Y|) + |\alpha_n^2 - \alpha_0^2|$$
$$+ 2|\alpha_0|\left\|\mathbb{Q}_n(Y \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(Y \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\right\|_{[a,b]} + \alpha_0^2 \left\|\mathbb{Q}_n(\mathbf{1}_{Z \leq (\cdot) \wedge \hat{\zeta}_n}) - \mathbb{P}(\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\right\|_{[a,b]}$$

and

$$\left\|\mathbb{Q}_n(|\tilde{\epsilon}_n| \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(|\epsilon| \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n})\right\|_{[a,b]} \leq$$
$$\left\|\mathbb{Q}_n(|Y - \alpha_0| \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(|Y - \alpha_0| \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\right\|_{[a,b]} + |\alpha_n - \alpha_0|.$$

and

$$\left\|\mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n})\right\|_{[a,b]} \leq \left\|\mathbb{Q}_n(Y \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(Y \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\right\|_{[a,b]}$$
$$+ |\alpha_n - \alpha_0| + |\alpha_0|\left\|\mathbb{Q}_n(\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}) - \mathbb{P}(\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_0})\right\|_{[a,b]}.$$

The other three cases follow from similar arguments. $\qquad\square$

**Lemma A.6** *Suppose that (I)-(IV) hold. Then,*

*(i)* $\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n})\right\|_{[a,b]} \xrightarrow{\mathbf{P}} 0,$

*(ii)* $\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n})\right\|_{[a,b]} \xrightarrow{\mathbf{P}} 0.$

*Also, these statements are true if $\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n}$ is replaced by any of $\mathbf{1}_{(\cdot) < Z \leq \zeta_n}$, $\mathbf{1}_{\zeta_n < Z \leq (\cdot)}$ or $\mathbf{1}_{Z > (\cdot) \vee \zeta_n}$.*

**Proof:** By the maximal inequality 3.1 from Kim and Pollard (1990) and Lemma A.4 we see that:

$$\mathbf{E}\left(\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n})\right\|_{[a,b]}\right) \leq \frac{J_{\mathscr{A}}(1)}{\sqrt{m_n}}\sqrt{\mathbb{Q}_n(\tilde{\epsilon}_n^2)}$$
$$\mathbf{E}\left(\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n})\right\|_{[a,b]}\right) \leq \frac{J_{\mathcal{F}}(1)}{\sqrt{m_n}}.$$

The lemma now follow directly as $\mathbb{Q}_n(\tilde{\epsilon}_n^2) \to \sigma^2$ (a consequence of Lemma A.5). The other statements are proven similarly. $\qquad\square$

### A.2.1 Proof of Proposition 3.1

Noting that $\tilde{\epsilon}_n = Y - \alpha_n \mathbf{1}_{Z \leq \zeta_n} - \beta_n \mathbf{1}_{Z > \zeta_n}$, we write

$$m_\theta(X) = -(\tilde{\epsilon}_n + \alpha_n - \alpha)^2 \mathbf{1}_{Z \leq \zeta_n \wedge \zeta} - (\tilde{\epsilon}_n + \beta_n - \alpha)^2 \mathbf{1}_{\zeta_n < Z \leq \zeta}$$
$$-(\tilde{\epsilon}_n + \alpha_n - \beta)^2 \mathbf{1}_{\zeta < Z \leq \zeta_n} - (\tilde{\epsilon}_n + \beta_n - \beta)^2 \mathbf{1}_{Z > \zeta_n \vee \zeta}, \quad (30)$$

and therefore

$$-\mathbb{P}_n^*(\tilde{\epsilon}_n^2) = M_n^*(\theta_n) \qquad \leq M_n^*(\theta_n^*)$$
$$\leq -\mathbb{P}_n^*[(\tilde{\epsilon}_n - \alpha_n^* + \alpha_n)^2 \mathbf{1}_{Z < a}] - \mathbb{P}_n^*[(\tilde{\epsilon}_n - \beta_n^* + \beta_n)^2 \mathbf{1}_{Z > b}].$$

Letting $\gamma_n^* = (\alpha_n^*, \beta_n^*)$, noticing that $M_n^*(\hat{\theta}_n) = -\mathbb{P}_n^*(\tilde{\epsilon}_n^2)$, and by rearranging the terms in the above inequality, we get

$$|\gamma_n^* - \gamma_n|^2 \mathbb{P}_n^*(Z < a) \wedge \mathbb{P}_n^*(Z > b) \leq \mathbb{P}_n^*\left(\tilde{\epsilon}_n^2 \mathbf{1}_{a \leq Z \leq b}\right)$$
$$+2|\gamma_n^* - \gamma_n|\left(|\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z < a}\right)| + |\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z > b}\right)|\right).$$

Consider $\mathbb{P}_n^*(Z < a)$. By $(ii)$ of Lemma A.6 we see that $|(\mathbb{P}_n^* - \mathbb{Q}_n)(Z < a)| \xrightarrow{\mathbf{P}} 0$ and by $(iv)$ of Lemma A.5 we can show that $|(\mathbb{Q}_n - \mathbb{P})(Z < a)| \to 0$. Thus, combining the two, we have $\mathbb{P}_n^*(Z < a) \xrightarrow{\mathbf{P}} \mathbb{P}(Z < a)$. Similarly, we can show that $\mathbb{P}_n^*(Z < a) \wedge \mathbb{P}_n^*(Z > b) \xrightarrow{\mathbf{P}} \mathbb{P}(Z < a) \wedge \mathbb{P}(Z > b) > 0$ and also that $|\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z < a}\right)| + |\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z > b}\right)| \xrightarrow{\mathbf{P}} 0$. Also, observe that $\mathbf{E}\left(\mathbb{P}_n^*(\tilde{\epsilon}_n^2)\right) = \mathbb{Q}_n(\tilde{\epsilon}_n^2) \to \sigma^2$, by assumptions (I)-(III) and so $\mathbb{P}_n^*(\tilde{\epsilon}_n^2)$ is bounded in $\mathbb{L}^1$. Hence, we can write

$$|\gamma_n^* - \gamma_n|^2 \leq O_{\mathbf{P}}(1) + |\gamma_n^* - \gamma_n|o_{\mathbf{P}}(1)$$

and therefore $|\gamma_n^* - \gamma_n| = O_{\mathbf{P}}(1)$ (and, consequently, $|\gamma_n^* - \gamma_0| = O_{\mathbf{P}}(1)$). We first rewrite $m_\theta(X)$ as follows:

$$m_\theta(X) = -\tilde{\epsilon}_n^2 - 2(\alpha_n - \alpha)\tilde{\epsilon}_n \mathbf{1}_{Z \leq \zeta \wedge \zeta_n} - (\alpha_n - \alpha)^2 \mathbf{1}_{Z \leq \zeta \wedge \zeta_n}$$
$$-2(\beta_n - \alpha)\tilde{\epsilon}_n \mathbf{1}_{\zeta_n < Z \leq \zeta} - (\beta_n - \alpha)^2 \mathbf{1}_{\zeta_n < Z \leq \zeta}$$
$$-2(\alpha_n - \beta)\tilde{\epsilon}_n \mathbf{1}_{\zeta < Z \leq \zeta_n} - (\alpha_n - \beta)^2 \mathbf{1}_{\zeta < Z \leq \zeta_n}$$
$$-2(\beta_n - \beta)\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta \vee \zeta_n} - (\beta_n - \beta)^2 \mathbf{1}_{Z > \zeta \vee \zeta_n}. \quad (31)$$

We can then decompose $M_n^*$ as in (31), and use Lemmas A.6 and A.5 and the fact that $\theta_n \to \theta_0$, to obtain

$$\left\| M_n^* + \mathbb{P}_n^*(\tilde{\epsilon}_n^2) - M_n - \mathbb{Q}_n(\tilde{\epsilon}_n^2) \right\|_K \xrightarrow{\mathbf{P}} 0.$$
$$\left\| M_n^* + \mathbb{P}_n^*(\tilde{\epsilon}_n^2) - M - \sigma^2 \right\|_K \xrightarrow{\mathbf{P}} 0$$

for every compact $K \subset \Theta$. But $\theta_0$ is also the unique maximizer of $M + \sigma^2$ and $|\gamma_n^* - \gamma_0| = O_{\mathbf{P}}(1)$. Therefore, the conditions of Corollary 3.2.3 (ii), page 287 of Van der Vaart and Wellner (1996), hold and we obtain that $\theta_n^* \xrightarrow{\mathbf{P}} \theta_0$ (and also that $\theta_n^* - \theta_n \xrightarrow{\mathbf{P}} 0$). $\qquad\square$

### A.2.2 Proof of Proposition 3.2

We will apply Theorem 3.4.1 of Van der Vaart and Wellner (1996) to prove the result. Let $d : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ be given by $d(\theta, \vartheta) = |(\theta_1, \theta_2) - (\vartheta_1, \vartheta_2)| + \sqrt{|\theta_3 - \vartheta_3|}$. Consider $\eta, \rho, L > 0$ as in (V) and a compact rectangle $K \subset \Theta$ such that $\{\theta \in \Theta : d(\theta, \theta_n) < \eta$ for some $n \in \mathbb{N}\} \subset K$. We can take $L$ large enough so $L > 1 \vee \sup\{|\theta_1 - \vartheta_2| \vee |\theta_2 - \vartheta_1| : \theta, \vartheta \in K\}$. Pick $n$ large enough so we can fix some $\delta \in (\frac{2\sqrt{2}}{m_n^{1/4}}, \eta)$. Then, taking also (I)-(IV) into account and possibly making $\eta$ smaller, we can find positive constants $c_1, c_2 > 0$ and $N \in \mathbb{N}$ such that for any $n \geq N$, we have (5), (6), (7) and the inequalities:

$$\inf_{d(\theta, \theta_n) < \delta} \left\{ |\alpha_n - \beta|^2 \wedge |\beta_n - \alpha|^2 \right\} > c_1,$$
$$\mathbb{Q}_n(Z \leq a) \wedge \mathbb{Q}_n(Z > b) > c_2.$$

Also, let $\mathbb{M}_n(\theta) := M_n^*(\theta) + \mathbb{P}_n^*(\tilde{\epsilon}_n^2)$ and $\mathcal{M}_n(\theta) := M_n(\theta) + \mathbb{Q}_n(\tilde{\epsilon}_n^2)$ for all $\theta \in \Theta$. Choose $n \geq N$ and $\theta \in \Theta$ with $\frac{\delta}{2} < d(\theta, \theta_n) < \delta$. Then, considering the properties of the constants just defined and the expression

$$
\begin{aligned}
\mathcal{M}_n(\theta) - \mathcal{M}_n(\theta_n) = {} & -2(\alpha_n - \alpha)\mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{Z \leq \zeta \wedge \zeta_n}) - (\alpha_n - \alpha)^2 \mathbb{Q}_n(\mathbf{1}_{Z \leq \zeta \wedge \zeta_n}) \\
& - 2(\beta_n - \alpha)\mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n < Z \leq \zeta}) - (\beta_n - \alpha)^2 \mathbb{Q}_n(\mathbf{1}_{\zeta_n < Z \leq \zeta}) \\
& - 2(\alpha_n - \beta)\mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{\zeta < Z \leq \zeta_n}) - (\alpha_n - \beta)^2 \mathbb{Q}_n(\mathbf{1}_{\zeta < Z \leq \zeta_n}) \\
& - 2(\beta_n - \beta)\mathbb{Q}_n(\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta \vee \zeta_n}) - (\beta_n - \beta)^2 \mathbb{Q}_n(\mathbf{1}_{Z > \zeta \vee \zeta_n}) \quad (32)
\end{aligned}
$$

it is seen that the sum of the 1st, 3rd, 5th, and 7th terms in (32) can be bounded

from above by $\frac{8L^2\delta}{\sqrt{m_n}}$. While we also have,

$$
\begin{aligned}
(\alpha_n - \alpha)^2 \mathbb{Q}_n(\mathbf{1}_{Z \leq \zeta \wedge \zeta_n}) &\geq c_2(\alpha_n - \alpha)^2, \\
(\beta_n - \beta)^2 \mathbb{Q}_n(\mathbf{1}_{Z > \zeta \vee \zeta_n}) &\geq c_2(\beta_n - \beta)^2, \\
(\beta_n - \alpha)^2 \mathbb{Q}_n(\mathbf{1}_{\zeta_n < Z \leq \zeta}) &\geq c_1\rho|\zeta - \zeta_n|, \quad \text{if } |\zeta - \zeta_n| \geq \frac{\delta^2}{8} > \frac{1}{\sqrt{m_n}}, \\
(\alpha_n - \beta)^2 \mathbb{Q}_n(\mathbf{1}_{\zeta < Z \leq \zeta_n}) &\geq c_1\rho|\zeta - \zeta_n|, \quad \text{if } |\zeta - \zeta_n| \geq \frac{\delta^2}{8} > \frac{1}{\sqrt{m_n}},
\end{aligned}
$$

and therefore, noting that either $(\alpha - \alpha_n)^2 + (\beta - \beta_n)^2 \geq \frac{\delta^2}{8}$ or $|\zeta - \zeta_n| \geq \frac{\delta^2}{8}$, letting $c = \frac{1}{16}c_2 \wedge (c_1\rho)$ and adding all the terms in the previous display, we get

$$
\sup_{\frac{\delta}{2} < d(\theta, \theta_n) < \delta} \{\mathcal{M}_n(\theta) - \mathcal{M}_n(\theta_n)\} \leq \frac{8L^2}{\sqrt{m_n}}\delta - 2c\delta^2 \quad \forall n \geq N.
$$

Hence, setting $\delta_n = \frac{8L^2}{c\sqrt{m_n}} \wedge \frac{2\sqrt{2}}{m_n^{1/4}}$ we get that

$$
\sup_{\frac{\delta}{2} < d(\theta, \theta_n) < \delta} \{\mathcal{M}_n(\theta) - \mathcal{M}_n(\theta_n)\} \leq -c\delta^2 \quad \forall \, \delta_n \leq \delta < \eta, \quad \forall n \geq N. \tag{33}
$$

Next we will show

$$
\sqrt{n}\mathbf{E}\left(\sup_{d(\theta, \theta_n) < \delta} \{|(\mathbb{M}_n - \mathcal{M}_n)(\theta) - (\mathbb{M}_n - \mathcal{M}_n)(\theta_n)|\}\right) \lesssim \frac{\sqrt{n}}{\sqrt{m_n}}\delta. \tag{34}
$$

Note that, using the expansion (31), $\mathbb{M}_n(\theta_n) = \mathcal{M}_n(\theta_n) = 0$. To control the term $(\mathbb{M}_n - \mathcal{M}_n)(\theta)$ observe that it admits a very similar expansion as (32) with the $\mathbb{Q}_n$ replaced by $(\mathbb{P}_n^* - \mathbb{Q}_n)$; in particular, we can write the difference $\mathbb{M}_n(\theta) - \mathcal{M}_n(\theta)$ (by re-arranging the terms) as

$$
\begin{aligned}
&- 2(\alpha_n - \alpha)(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{Z \leq \zeta \wedge \zeta_n}) - 2(\beta_n - \beta)(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta \vee \zeta_n}) \\
&-2(\beta_n - \alpha)(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n < Z \leq \zeta}) - 2(\alpha_n - \beta)(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{\zeta < Z \leq \zeta_n}) \\
&-(\alpha_n - \alpha)^2(\mathbb{P}_n^* - \mathbb{Q}_n)(\mathbf{1}_{Z \leq \zeta \wedge \zeta_n}) - (\beta_n - \beta)^2(\mathbb{P}_n^* - \mathbb{Q}_n)(\mathbf{1}_{Z > \zeta \vee \zeta_n}) \\
&-(\alpha_n - \beta)^2(\mathbb{P}_n^* - \mathbb{Q}_n)(\mathbf{1}_{\zeta < Z \leq \zeta_n}) - (\beta_n - \alpha)^2(\mathbb{P}_n^* - \mathbb{Q}_n)(\mathbf{1}_{\zeta_n < Z \leq \zeta}). \tag{35}
\end{aligned}
$$

Each of these terms can be controlled by using Lemma A.4 as

$$
\begin{aligned}
\mathbf{E}\left(\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{Z \leq (\cdot) \wedge \zeta_n})\right\|_{[a,b]}\right) &\leq \frac{J_{\mathscr{A}}(1)}{\sqrt{m_n}}\sqrt{\mathbb{Q}_n(\tilde{\epsilon}_n^2)} \\
\mathbf{E}\left(\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{(\cdot) < Z \leq \zeta_n})\right\|_{|\zeta - \zeta_n| < \delta^2}\right) &\leq \frac{J_{\mathscr{A}}(1)}{\sqrt{m_n}}\sqrt{\mathbb{Q}_n(\tilde{\epsilon}_n^2 \mathbf{1}_{\zeta_n - \delta^2 < Z \leq \zeta_n + \delta^2})}.
\end{aligned}
$$

Lemma A.5 implies that $\mathbb{Q}_n(\tilde{\epsilon}_n^2 \mathbf{1}_{\zeta_n - \delta^2 < Z \le \zeta_n + \delta^2}) \to \sigma^2 \mathbb{P}(\zeta_0 - \delta^2 < Z \le \zeta_0 + \delta^2) = \sigma^2\{2f(\zeta_0)\delta^2 + o(\delta^2)\}$. Hence, there is a constant $R > 0$ such that the right side of the above equations are bounded by $R/\sqrt{m_n}$ and $R\sqrt{\delta^2 + o(\delta^2)}/\sqrt{m_n}$. Using similar arguments, we can in fact make $R$ large enough so that the following inequalities hold too

$$\mathbf{E}\left(\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{Z > (\cdot) \vee \zeta_n})\right\|_{[a,b]}\right) \le \frac{R}{\sqrt{m_n}} \tag{36}$$

$$\mathbf{E}\left(\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n < Z \le (\cdot)})\right\|_{|\zeta - \zeta_n| < \delta^2}\right) \le \frac{R}{\sqrt{m_n}} \sqrt{\delta^2 + o(\delta^2)}. \tag{37}$$

We also assume that $R > J_{\mathcal{F}}(1)$. Using (36), (37), the discussion preceding the display, and grouping two consecutive terms at a time in the expansion (35), it is easily seen that

$$\sqrt{n}\mathbf{E}\left(\sup_{d(\theta,\theta_n) < \delta}\{|(\mathbb{M}_n - \mathcal{M}_n)(\theta) - (\mathbb{M}_n - \mathcal{M}_n)(\theta_n)|\}\right) \lesssim \frac{4R\sqrt{n}}{\sqrt{m_n}}\delta$$

$$+ \frac{4RL\sqrt{n}}{\sqrt{m_n}}\sqrt{\delta^2 + o(\delta^2)} + \frac{2R\sqrt{n}}{\sqrt{m_n}}\delta^2 + \frac{2RL^2 f(\zeta_0)\sqrt{n}}{\sqrt{m_n}}(\delta^2 + o(\delta^2)).$$

Thus by taking $\eta > 0$ small enough we can show that (34) holds for every $n \ge N$ and any $\delta \in [\delta_n, \eta)$, with $\delta_n$ and $N$ defined as in (33). Defining $\phi_n(\delta) = \frac{\sqrt{n}}{\sqrt{m_n}}\delta$ and $r_n = \sqrt{m_n}$, the hypotheses of Theorem 3.4.1 of Van der Vaart and Wellner (1996) are satisfied (note that Proposition 3.1 implies that $d(\theta_n, \theta_n^*) \xrightarrow{\mathbf{P}} 0$). Therefore, $r_n d(\theta_n, \theta_n^*) = \sqrt{m_n(\alpha_n^* - \alpha_n)^2 + m_n(\beta_n^* - \beta_n)^2} + \sqrt{m_n}|\zeta_n^* - \zeta_n| = O_{\mathbf{P}}(1)$. $\square$

### A.2.3   Proof of Lemma 3.1

Let $\eta > 0$ be an upper bound for the norm of the elements in $K$. The maximal inequality from Kim and Pollard (1990) and Lemma A.4 imply

$$\sqrt{m_n}\mathbf{E}\left(\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n + \frac{(\cdot)}{m_n} < Z \le \zeta_n})\right\|_K\right) \le J_{\mathscr{A}}(1)\sqrt{\mathbb{Q}_n(\tilde{\epsilon}_n^2 \mathbf{1}_{\zeta_n - \frac{\eta}{m_n} < Z \le \zeta_n})}$$

$$\sqrt{m_n}\mathbf{E}\left(\left\|(\mathbb{P}_n^* - \mathbb{Q}_n)(\mathbf{1}_{\zeta_n + \frac{(\cdot)}{m_n} < Z \le \zeta_n})\right\|_K\right) \le J_{\mathcal{F}}(1)\sqrt{\mathbb{Q}_n(\mathbf{1}_{\zeta_n - \frac{\eta}{m_n} < Z \le \zeta_n})}.$$

By $(i)$ and $(iv)$ of Lemma A.5 applied with $\mathbf{1}_{Z \le (\cdot) \wedge \zeta_n}$ in place of $\mathbf{1}_{(\cdot) < Z \le \zeta_n}$, we see that the righthand side of both the above inequalities go to zero. On the other hand, using (8) and (9) it is easy to see that both $\sqrt{m_n}\|\mathbb{Q}_n(\tilde{\epsilon}_n^2 \mathbf{1}_{\zeta_n + \frac{(\cdot)}{m_n} < Z \le \zeta_n})\|_K$ and $\sqrt{m_n}\|\mathbb{Q}_n(\mathbf{1}_{\zeta_n + \frac{(\cdot)}{m_n} < Z \le \zeta_n})\|_K$ converge to zero. Now, note that

38

$\sqrt{m_n}\left\|\mathbb{P}_n^*\left(\tilde{\epsilon}_n\mathbf{1}_{\zeta_n+\frac{(\cdot)}{m_n}<Z\leq\zeta_n}\right)\right\|_K$ is bounded by

$$\sqrt{m_n}\left\|(\mathbb{P}_n^*-\mathbb{Q}_n)(\tilde{\epsilon}_n\mathbf{1}_{\zeta_n+\frac{(\cdot)}{m_n}<Z\leq\zeta_n})\right\|_K + \sqrt{m_n}\left\|\mathbb{Q}_n\left(|\tilde{\epsilon}_n|\mathbf{1}_{\zeta_n+\frac{(\cdot)}{m_n}<Z\leq\zeta_n}\right)\right\|_K$$

and thus $\sqrt{m_n}\left\|\mathbb{P}_n^*\left(\tilde{\epsilon}_n\mathbf{1}_{\zeta_n+\frac{(\cdot)}{m_n}<Z\leq\zeta_n}\right)\right\|_K \xrightarrow{\mathbb{L}_1} 0$. Similarly we can bound $\sqrt{m_n}\left\|\mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n+\frac{(\cdot)}{m_n}<Z\leq\zeta_n}\right)\right\|_K$ and show that it converges to zero in mean. Finally, from the expressions

$$
\begin{aligned}
A_n^*(h_1) - \hat{A}_n(h_1, h_3) &= 2h_1\sqrt{m_n}\mathbb{P}_n^*\left(\tilde{\epsilon}_n\mathbf{1}_{\zeta_n+\frac{h_3}{m_n}<Z\leq\zeta_n}\right) - h_1^2\mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n+\frac{(h_3)}{m_n}<Z\leq\zeta_n}\right), \\
C_n^*(h_3) - \hat{C}_n(h_2, h_3) &= 2h_2\sqrt{m_n}\mathbb{P}_n^*\left(\tilde{\epsilon}_n\mathbf{1}_{\zeta_n+\frac{h_3}{m_n}<Z\leq\zeta_n}\right) \\
&\quad - \left(2h_2\sqrt{m_n}(\alpha_n-\beta_n) - h_2^2\right)\mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n+\frac{h_3}{m_n}<Z\leq\zeta_n}\right)
\end{aligned}
$$

we get that $\left\|A_n^* - \hat{A}_n\right\|_K \xrightarrow{\mathbb{L}_1} 0$ and $\left\|C_n^* - \hat{C}_n\right\|_K \xrightarrow{\mathbb{L}_1} 0$. With completely analogous arguments, it is seen that $\left\|B_n^* - \hat{B}_n\right\|_K \xrightarrow{\mathbb{L}_1} 0$ and $\left\|D_n^* - \hat{D}_n\right\|_K \xrightarrow{\mathbb{L}_1} 0$ as well. Observing that $\hat{E}_n = \hat{A}_n + \hat{B}_n + \hat{C}_n + \hat{D}_n - \mathbb{P}_n^*(\tilde{\epsilon}_n^2)$ completes the proof of the result. $\square$

### A.2.4  Proof of Lemma 3.2

It suffices to show that each of the components of $(\Xi_n)_{n=1}^\infty$ is tight. Write $\tilde{\epsilon}_{n,j} = \tilde{\epsilon}_n(Z_{n,j}, Y_{n,j})$ and let

$$
\begin{aligned}
r_n &= m_n\mathbb{Q}_n\left(e^{i\frac{\xi}{\sqrt{m_n}}\tilde{\epsilon}_n\mathbf{1}_{Z\leq\zeta_n}} - 1 - i\frac{\xi}{m_n}\sqrt{m_n}\tilde{\epsilon}_n\mathbf{1}_{Z\leq\zeta_n} + \frac{\xi^2}{2m_n}\tilde{\epsilon}_n^2\mathbf{1}_{Z\leq\zeta_n}\right) \\
&\leq \frac{m_n^{-1/2}\xi^3\mathbb{Q}_n|\tilde{\epsilon}_n|^3}{6}.
\end{aligned}
$$

Then, assumption (VIII) implies that $r_n \to 0$ as $n \to \infty$. Since the characteristic function of $\sqrt{m_n}\mathbb{P}_n^*(\tilde{\epsilon}_n\mathbf{1}_{Z\leq\zeta_n})$ is given by

$$\mathbf{E}\left(e^{i\xi\sqrt{m_n}\mathbb{P}_n^*(\tilde{\epsilon}_n\mathbf{1}_{Z\leq\zeta_n})}\right) = \left(1 + i\frac{\xi}{\sqrt{m_n}}\mathbb{Q}_n\left(\tilde{\epsilon}_n\mathbf{1}_{Z\leq\zeta_n}\right) - \frac{\xi^2}{2m_n}\mathbb{Q}_n\left(\tilde{\epsilon}_n^2\mathbf{1}_{Z\leq\zeta_n}\right) + \frac{r_n}{m_n}\right)^{m_n}$$

taking the limit as $n \to \infty$ we can conclude that $\sqrt{m_n}\mathbb{P}_n^*(\tilde{\epsilon}_n\mathbf{1}_{Z\leq\zeta_n}) \rightsquigarrow N(0, \mathbb{P}(Z \leq \zeta_0)\sigma^2)$ by using (VII) and the fact that $(1 + \beta_n/n)^n \to e^\beta$ if $\beta_n \to \beta$. With similar arguments, it is seen that $\sqrt{m_n}\mathbb{P}_n^*(\tilde{\epsilon}_n\mathbf{1}_{Z>\zeta_n}) \rightsquigarrow N(0, \mathbb{P}(Z > \zeta_0)\sigma^2)$, so the first two components of the random vector of interest are uniformly tight.

Consider now the processes $\Gamma_n(t) = m_n \mathbb{P}_n^* (\mathbf{1}_{\zeta_n < Z \leq \zeta_n + \frac{t}{m_n}})$ and $\Psi_n(t) = m_n \mathbb{P}_n^* (\tilde{\epsilon}_n \mathbf{1}_{\zeta_n < Z \leq \zeta_n + \frac{t}{m_n}})$. For any process $\Psi \in \mathcal{D}_I$, $I \subset \mathbb{R}$ compact interval, $\delta > 0$, we write

$$w_\Psi''(\delta) = \sup\left\{ |\Psi(t_1) - \Psi(t)| \wedge |\Psi(t_2) - \Psi(t)| \right\}$$

where the supremum is taken over all $t_1 \leq t \leq t_2 \in I$ with $0 \leq t_2 - t_1 \leq \delta$. Also, for any $A \subset I$, define $w_\Psi(A) = \sup_{s,t \in A} \{|\Psi(t) - \Psi(s)|\}$. This agrees with the notation defined in Chapter 14 of Billingsley (1968). Let $\eta > 0$ be an upper bound for the absolute values of the elements of $I$, consider any $\rho > 0$, and define the numbers $a_\Psi^\rho$ and $a_\Gamma^\rho$ by,

$$a_\Psi^\rho = \frac{1}{\rho} \sup_{n \in \mathbb{N}} \left\{ m_n \mathbb{Q}_n \left( |\tilde{\epsilon}_n| \mathbf{1}_{\zeta_n < Z \leq \zeta_n + \frac{\eta}{m_n}} \right) \right\}$$

$$a_\Gamma^\rho = \frac{1}{\rho} \sup_{n \in \mathbb{N}} \left\{ m_n \mathbb{Q}_n \left( \mathbf{1}_{\zeta_n < Z \leq \zeta_n + \frac{\eta}{m_n}} \right) \right\}.$$

Then, using Markov's inequality,

$$\varlimsup_{n \to \infty} \mathbf{P} \left( \sup_{t \in I} \{|\Psi_n(t)|\} > a_\Psi^\rho \right) \leq \rho \tag{38}$$

$$\varlimsup_{n \to \infty} \mathbf{P} \left( \sup_{t \in I} \{|\Gamma_n(t)|\} > a_\Gamma^\rho \right) \leq \rho. \tag{39}$$

Now, let $\rho, \gamma > 0$ be any pair of positive numbers and assume that $I = [a, b]$. Then, choose $\delta < \frac{\gamma}{8|b-a|f(\zeta_0)^2} \wedge \frac{|b-a|}{4} \wedge \frac{1}{f(\zeta_0)}$ so there is an integer $N \geq 2$ such that $\delta < \frac{|b-a|}{N} < 2\delta$. Define $s_j = a + \frac{j}{N}(b - a)$ and consider the partition $\{a = s_0 < s_1 < \ldots < s_N = b\}$ of $I$. Notice that if $\Psi$ is a step function on $I$, for $w_\Psi''(\delta)$ to be positive, we need at least two jumps in an interval of size at most $\delta$. Then, the probability that at least two jumps of the process $\Psi_n$ happens on any interval $(s_{j-2}, s_j]$ is bounded from above by

$$a_{j,m_n} := \mathbf{P} \left( \bigcup_{1 \leq k < l \leq m_n} \left[ m_n(Z_{n,k} - \zeta_n), m_n(Z_{n,l} - \zeta_n) \in (s_{j-2}, s_j] \right] \right)$$

$$\leq \frac{m_n^2}{2} \mathbb{Q}_n \left( \zeta_n + \frac{s_{j-2}}{m_n} < Z \leq \zeta_n + \frac{s_j}{m_n} \right)^2$$

and hence the limit superior of the probability that either $\Psi_n$ or $\Gamma_n$ has two jumps in any interval of the form $(s_{j-2}, s_j]$ is bounded from above by $2|b-a|^2 f(\zeta_0)^2 / N^2$ by (VI). Therefore, the probability that at least two jumps happen in any interval

of size at most $\delta$ is asymptotically bounded from above by

$$\sum_{i=2}^{N} a_{j,m_n} \le \sum_{i=2}^{N} 2|b-a|^2 f(\zeta_0)^2/N^2 \le 4(N-1)f(\zeta_0)^2|b-a|\delta/N \le \gamma.$$

Thus,

$$\varlimsup_{n\to\infty} \mathbf{P}\left(w_{\Psi_n}''(\delta) > \rho\right) \; < \; \gamma \tag{40}$$

The exact same argument can be used to show that

$$\varlimsup_{n\to\infty} \mathbf{P}\left(w_{\Gamma_n}''(\delta) > \rho\right) \; < \; \gamma. \tag{41}$$

Now, note that

$$\mathbf{P}\left(w_{\Psi_n}([a, a+\delta)) > \rho\right) \; \le \; \mathbf{P}\left(\bigcup_{j=1}^{m_n} m_n(Z_{n,j} - \zeta_n) \in [a, a+\delta) > \rho\right)$$

$$\le \; m_n \mathbb{Q}_n\left(\zeta_n + \frac{a}{m_n} < Z \le \zeta_n + \frac{a+\delta}{m_n}\right)$$

which implies that

$$\varlimsup_{n\to\infty} \mathbf{P}\left(w_{\Psi_n}([a, a+\delta)) > \rho\right) \le \delta f(\zeta_0) < \gamma. \tag{42}$$

A similar analysis leads to the following bounds

$$\varlimsup_{n\to\infty} \mathbf{P}\left(w_{\Psi_n}([b-\delta, b)) > \rho\right) \; < \; \gamma \tag{43}$$

$$\varlimsup_{n\to\infty} \mathbf{P}\left(w_{\Gamma_n}([a, a+\delta)) > \rho\right) \; < \; \gamma \tag{44}$$

$$\varlimsup_{n\to\infty} \mathbf{P}\left(w_{\Gamma_n}([b-\delta, b)) > \rho\right) \; < \; \gamma. \tag{45}$$

Putting together (38), (39), (40), (41), (42), (43), (44) and (45) and using Theorem 15.3 of Billingsley (1968) we obtain that both sequences $(\Psi_n)_{n=1}^\infty$ and $(\Gamma_n)_{n=1}^\infty$ are uniformly tight in $\tilde{\mathcal{D}}_I$. Similar arguments show the tightness of the third and fourth components of the process. Therefore, $(\Xi_n)_{n=1}^\infty$ is uniformly tight. The uniform tightness of $(E_n^*)_{n=1}^\infty$ now follows from the fact that $(\Xi_n)_{n=1}^\infty$ is uniformly tight and $E_n^*$ is a continuous function of $\Xi_n$. $\qquad\square$

### A.2.5 Proof of Lemma 3.3

In view of Lemma 3.2, to show $(i)$ it suffices to show convergence of the finite dimensional distributions. To this end, consider the real numbers $t_{-N_-} < \ldots <$

$t_{-1} < 0 = t_0 < t_1 < \ldots < t_{N_+}$ and the linear combination

$$
\begin{aligned}
W_n &= \mu\sqrt{m_n}\mathbb{P}_n^*(\tilde{\epsilon}_n \mathbf{1}_{Z \le \zeta_n}) + \lambda\sqrt{m_n}\mathbb{P}_n^*(\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta_n}) \\
&\quad + \sum_{j=1}^{N_-}\left\{\xi_{-j}m_n\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n + \frac{t_{-j}}{m_n} < Z \le \zeta_n}\right) + \eta_{-j}m_n\mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n + \frac{t_{-j}}{m_n} < Z \le \zeta_n}\right)\right\} \\
&\quad + \sum_{j=1}^{N_+}\left\{\xi_j m_n\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n < Z \le \zeta_n + \frac{t_j}{m_n}}\right) + \eta_j m_n\mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n < Z \le \zeta_n + \frac{t_j}{m_n}}\right)\right\} \quad (46)
\end{aligned}
$$

where $\mu$, $\lambda$ and the $\xi_j$'s and the $\eta_j$'s are arbitrary real numbers. Now, set $\xi_0 = \eta_0 = 0$ and define

$$
\mu_{\pm j} = \sum_{k=j}^{N_\pm} \eta_{\pm k} \quad \text{and} \quad \lambda_{\pm j} = \sum_{k=j}^{N_\pm} \xi_{\pm k}. \tag{47}
$$

Then grouping terms appropriately we can rewrite $W_n$ as

$$
\begin{aligned}
W_n &= \mu\sqrt{m_n}\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z \le \zeta_n + \frac{t_{-N_-}}{m_n}}\right) + \lambda\sqrt{m_n}\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{Z > \zeta_n + \frac{t_{N_+}}{m_n}}\right) \\
&\quad + \sum_{j=1}^{N_-}(\lambda_{-j}m_n + \mu\sqrt{m_n})\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n + \frac{t_{-j}}{m_n} < Z \le \zeta_n + \frac{t_{-j+1}}{m_n}}\right) \\
&\quad + \sum_{j=1}^{N_-}\mu_{-j}m_n\mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n + \frac{t_{-j}}{m_n} < Z \le \zeta_n + \frac{t_{-j+1}}{m_n}}\right) \\
&\quad + \sum_{j=1}^{N_+}(\lambda_j m_n + \lambda\sqrt{m_n})\mathbb{P}_n^*\left(\tilde{\epsilon}_n \mathbf{1}_{\zeta_n + \frac{t_{j-1}}{m_n} < Z \le \zeta_n + \frac{t_j}{m_n}}\right) \\
&\quad + \sum_{j=1}^{N_+}\mu_j m_n\mathbb{P}_n^*\left(\mathbf{1}_{\zeta_n + \frac{t_{j-1}}{m_n} < Z \le \zeta_n + \frac{t_j}{m_n}}\right).
\end{aligned}
$$

Using the independence of $X_{n,1}, \ldots, X_{n,m_n}$, the characteristic function of $W_n$ is

$$
\begin{aligned}
\mathbf{E}\left(e^{isW_n}\right) &= \Bigg[1 + \sum_{j=1}^{N_-}\mathbb{Q}_n\left((e^{is(\frac{\mu}{\sqrt{m_n}} + \lambda_{-j})\tilde{\epsilon}_n + is\mu_{-j}} - 1)\mathbf{1}_{\zeta_n + \frac{t_{-j}}{m_n} < Z \le \zeta_n + \frac{t_{-j+1}}{m_n}}\right) \\
&\quad + \mathbb{Q}_n\left((e^{i\frac{s\mu}{\sqrt{m_n}}\tilde{\epsilon}_n} - 1)\mathbf{1}_{Z \le \zeta_n + \frac{t_{-N_-}}{m_n}}\right) + \mathbb{Q}_n\left((e^{i\frac{s\lambda}{\sqrt{m_n}}\tilde{\epsilon}_n} - 1)\mathbf{1}_{Z > \zeta_n + \frac{t_{N_+}}{m_n}}\right) \\
&\quad + \sum_{j=1}^{N_+}\mathbb{Q}_n\left((e^{is(\frac{\lambda}{\sqrt{m_n}} + \lambda_j)\tilde{\epsilon}_n + is\mu_j} - 1)\mathbf{1}_{\zeta_n + \frac{t_{j-1}}{m_n} < Z \le \zeta_n + \frac{t_j}{m_n}}\right)\Bigg]^{m_n}. \quad (48)
\end{aligned}
$$

Let $r_n$ be given by

$$
r_n = m_n\mathbb{Q}_n\left[\left(e^{i\frac{s\mu}{\sqrt{m_n}}\tilde{\epsilon}_n} - 1 - i\frac{s\mu}{\sqrt{m_n}}\tilde{\epsilon}_n + \frac{s^2\mu^2}{2m_n}\tilde{\epsilon}_n^2\right)\mathbf{1}_{Z \le \zeta_n + \frac{t_{-N_-}}{m_n}}\right] \le \frac{s^3\mathbb{Q}_n|\tilde{\epsilon}_n^3|}{6\sqrt{m_n}}.
$$

Condition (VIII) now implies that $r_n = o(1)$. But note that

$$\mathbb{Q}_n\left((e^{i\frac{s\mu}{\sqrt{m_n}}\tilde{\epsilon}_n} - 1)\mathbf{1}_{Z\leq\zeta_n+\frac{t_{-N_-}}{m_n}}\right) = i\frac{s\mu}{m_n}\sqrt{m_n}\mathbb{Q}_n\left(\tilde{\epsilon}\mathbf{1}_{Z\leq\zeta_n+\frac{t_{-N_-}}{m_n}}\right)$$
$$- \frac{s^2\mu^2}{2m_n}\mathbb{Q}_n\left(\tilde{\epsilon}_n^2\mathbf{1}_{Z\leq\zeta_n+\frac{t_{-N_-}}{m_n}}\right) + \frac{r_n}{m_n}$$

and so $(i)$ of Lemma A.5 together with condition (VII) and (8) imply that

$$m_n\mathbb{Q}_n\left((e^{i\frac{s\mu}{\sqrt{m_n}}\tilde{\epsilon}_n} - 1)\mathbf{1}_{Z\leq\zeta_n+\frac{t_{-N_-}}{m_n}}\right) = -\frac{s^2\mu^2}{2}\sigma^2\mathbb{P}(Z \leq \zeta_0) + o(1). \qquad (49)$$

Following a completely analogous argument one can show that

$$m_n\mathbb{Q}_n\left(\left(e^{i\frac{s\lambda}{\sqrt{m_n}}\tilde{\epsilon}_n} - 1\right)\mathbf{1}_{Z>\zeta_n+\frac{t_{N_+}}{m_n}}\right) = -\frac{s^2\lambda^2}{2}\sigma^2\mathbb{P}(Z > \zeta_0) + o(1). \qquad (50)$$

Now, take $1 \leq j \leq N_+$, and observe that equation (8) implies

$$m_n\left|\mathbb{Q}_n\left((e^{is(\frac{\lambda}{\sqrt{m_n}}+\lambda_j)\tilde{\epsilon}_n+is\mu_j} - e^{is\lambda_j\tilde{\epsilon}_n+is\mu_j})\mathbf{1}_{\zeta_n+\frac{t_{j-1}}{m_n}<Z\leq\zeta_n+\frac{t_j}{m_n}}\right)\right|$$
$$\leq |s\lambda|\sqrt{m_n}\mathbb{Q}_n\left(|\tilde{\epsilon}_n|\mathbf{1}_{\zeta_n+\frac{t_{j-1}}{m_n}<Z\leq\zeta_n+\frac{t_j}{m_n}}\right) \to 0.$$

Using (VI) we can write

$$m_n\mathbb{Q}_n\left((e^{is(\frac{\lambda}{\sqrt{m_n}}+\lambda_j)\tilde{\epsilon}_n+is\mu_j} - 1)\mathbf{1}_{\zeta_n+\frac{t_{j-1}}{m_n}<Z\leq\zeta_n+\frac{t_j}{m_n}}\right)$$
$$= (\varphi(s\lambda_j)e^{is\mu_j} - 1)f(\zeta_0)(t_j - t_{j-1}) + o(1)$$

where $\varphi$ is the characteristic function of $\epsilon$ (under $\mathbb{P}$). Thus,

$$m_n\sum_{j=1}^{N_+}\mathbb{Q}_n\left((e^{is(\frac{\lambda}{\sqrt{m_n}}+\lambda_j)\tilde{\epsilon}_n+is\mu_j} - 1)\mathbf{1}_{\zeta_n+\frac{t_{j-1}}{m_n}<Z\leq\zeta_n+\frac{t_j}{m_n}}\right)$$
$$= \sum_{j=1}^{N_+}(t_j - t_{j-1})f(\zeta_0)(\varphi(s\lambda_j)e^{is\mu_j} - 1) + o(1). \qquad (51)$$

Similarly, one can prove that

$$m_n\sum_{j=1}^{N_-}\mathbb{Q}_n\left((e^{is(\frac{\mu}{\sqrt{m_n}}+\lambda_{-j})\tilde{\epsilon}_n+is\mu_{-j}} - 1)\mathbf{1}_{\zeta_n+\frac{t_{-j}}{m_n}<Z\leq\zeta_n+\frac{t_{-j+1}}{m_n}}\right)$$
$$= \sum_{j=1}^{N_-}(t_{-j+1} - t_{-j})f(\zeta_0)(\varphi(s\lambda_{-j})e^{is\mu_{-j}} - 1) + o(1). \qquad (52)$$

43

So putting (46), (47), (48), (49), (50), (51) and (52) together we see that,

$$
\mathbf{E}\left(e^{isW_n}\right) \rightarrow \exp\left[\sum_{j=1}^{N_-} f(\zeta_0)(t_{-j+1} - t_{-j})\left\{\varphi\left(s(\sum_{k=j}^{N_-} \xi_{-k})\right)e^{is\sum_{k=j}^{N_-}\eta_{-k}} - 1\right\}\right.
$$

$$
- \frac{s^2\mu^2\sigma^2}{2}\mathbb{P}(Z \le \zeta_0) - \frac{s^2\lambda^2\sigma^2}{2}\mathbb{P}(Z > \zeta_0)
$$

$$
\left. + \sum_{j=1}^{N_+} f(\zeta_0)(t_j - t_{j-1})\left\{\varphi\left(s(\sum_{k=j}^{N_+} \xi_k)\right)e^{is\left(\sum_{k=j}^{N_+}\eta_k\right)} - 1\right\}\right]. \quad (53)
$$

But the right-hand side of (53) is precisely $\mathbf{E}\left(e^{isW}\right)$ where, with the notation of (11), $W$ is given by

$$
W = \mu\mathbf{Z}_1 + \lambda\mathbf{Z}_2 + \sum_{k=1}^{N_-}\left(\xi_{-k}\sum_{0<j\le\nu_1(-t_{-k})}v_k\mathbf{1}_{t_{-k}<0} + \eta_{-k}\nu_1(-t_{-k})\mathbf{1}_{t_{-k}<0}\right)
$$

$$
+ \sum_{k=1}^{N_+}\left(\xi_k\sum_{0<j\le\nu_2(t_k)}u_k\mathbf{1}_{t_k\ge0} + \eta_k\nu_2(t_k)\mathbf{1}_{t_k\ge0}\right)
$$

and thus $W_n \rightsquigarrow W$. From the fact that $\mu$, $\lambda$, the $\xi_j$'s and the $\eta_j$'s were arbitrarily chosen, by the Cramer-Wold device

$$
\left(\Xi_n(t_{-N_-}),\ldots,\Xi_n(t_{-1}),\Xi_n(t_1),\ldots,\Xi_n(t_{N_+})\right)' \rightsquigarrow \left(\Xi(t_{-N_-}),\ldots,\Xi(t_{-1}),\Xi(t_1),\ldots,\Xi(t_{N_+})\right)'.
$$

This gives the convergence of the finite dimensional distributions, proving $(i)$. An application of the continuous mapping theorem shows that $(i)$ implies $(ii)$. Further, Lemma 3.1 and $(ii)$ now imply $(iii)$. $\qquad\square$

### A.2.6 Proof of Lemma 3.4

Every sample path of $E^* = E^*(h_1, h_2, h_3)$ can be written as

$$
2h_1\mathbf{Z}_1 - h_1^2\mathbb{P}(Z \le \zeta_0) + 2h_2\mathbf{Z}_2 - h_2^2\mathbb{P}(Z > \zeta_0) + \mathbf{1}_{h_3<0}2(\alpha_0 - \beta_0)\sum_{j=1}^{\nu_1(-h_3)}v_j
$$

$$
-(\alpha_0 - \beta_0)^2\nu_1(-h_3)\mathbf{1}_{h_3<0} + \mathbf{1}_{h_3\ge0}2(\beta_0 - \alpha_0)\sum_{j=1}^{\nu_2(h_3)}u_j - \mathbf{1}_{h_3\ge0}(\alpha_0 - \beta_0)^2\nu_2(h_3).
$$

From this last expression it is obvious that for any fixed $h_3$, the $E^*(\cdot, \cdot, h_3)$ gets maximized at $\phi_1^* = \mathbf{Z}_1/\mathbb{P}(Z \le \zeta_0)$ and $\phi_2^* = \mathbf{Z}_2/\mathbb{P}(Z > \zeta_0)$. The independence of the three co-ordinates follows from the fact that $\phi_1^*$ depends only on $\mathbf{Z}_1$, $\phi_2^*$

depends only on $\mathbf{Z}_2$, and $\phi_3^*$ depends only on $\mathbf{u}$, $\mathbf{v}$, $\nu_1$ and $\nu_2$. Since $E^*$ is piecewise constant in the third argument $h_3$, to complete the proof it is enough to show that $E^*(\phi_1^*, \phi_2^*, h_3) \to -\infty$ as $|h_3| \to \infty$. But this follows from the law of the iterated logarithm (applied to the random walks defined by the $v_i$'s and $u_i$'s) together with the fact that $\nu_1(t) \wedge \nu_2(t) \xrightarrow{a.s.} \infty$ as $t \to \infty$. Note that $\sum_{j=1}^{\nu_1(-h_3)} v_j$ is of order $O(\sqrt{\nu_1 \log \log \nu_1})$ a.s. as $h_3 \to \infty$. $\qquad\square$

### A.2.7 Proof of Proposition 3.3

Lemma 3.4 and the fact that the $u_i$'s and the $v_i$'s come from a continuous distribution, show that $(E^*, J^*)$ satisfy the hypotheses of Lemma A.3, and in particular that (29) holds. Moreover, Proposition 3.2 shows that the sequence $(\sqrt{m_n}(\alpha_n^* - \alpha_n), \sqrt{m_n}(\beta_n^* - \beta_n), m_n(\zeta_n^* - \zeta_n))$ is tight. Now, consider $C \in \mathbb{N}$ and let $\phi_n$, $\phi_{n,C}$ and $\phi_C$ be the smallest maximizers of $\hat{E}_n$, $\hat{E}_n|_{[-C,C]^3}$ and $E^*|_{[-C,C]^3}$. To prove the result, we will apply Lemma A.3 and Lemma 3.3 of Lan et al. (2009). Using the notation of the latter, set $\epsilon = \frac{1}{C}$, $W_{n\epsilon} = \phi_{n,C}$, $W_\epsilon = \phi_C$, $W_n = \phi_n$ and $W = \phi^*$. From Proposition 3.2 we see that $\lim_{\epsilon \to 0} \varlimsup_{n \to \infty} \mathbf{P}(W_{n\epsilon} \neq W_n) = 0$. Lemma 3.4 implies that $\lim_{\epsilon \to 0} \mathbf{P}(W_\epsilon \neq W) = 0$. Finally, Lemma A.3 and an application of Skorohod's Representation Theorem (see Theorem 1.8, page 102 of Ethier and Kurtz (2005)) show that $W_{n\epsilon} \rightsquigarrow W_\epsilon$ and hence, from Lemma 3.3 of Lan et al. (2009), we conclude that $\phi_n \rightsquigarrow \phi^*$. $\qquad\square$

### A.2.8 Proof of Lemma 4.1

We expand $m_\theta(X)$ as in (30) but with $\epsilon = Y - \alpha_0 \mathbf{1}_{Z \leq \zeta_0} - \beta_0 \mathbf{1}_{Z > \zeta_0}$ in place of $\tilde{\epsilon}_n$ to get

$$
\begin{aligned}
m_\theta(X) &= -(\epsilon + \alpha_0 - \alpha)^2 \mathbf{1}_{Z \leq \zeta_0 \wedge \zeta} - (\epsilon + \beta_0 - \alpha)^2 \mathbf{1}_{\zeta_0 < Z \leq \zeta} \\
&\quad -(\epsilon + \alpha_0 - \beta)^2 \mathbf{1}_{\zeta < Z \leq \zeta_0} - (\epsilon + \beta_0 - \beta)^2 \mathbf{1}_{Z > \zeta_0 \vee \zeta}.
\end{aligned} \tag{54}
$$

Letting $\hat{\gamma}_n = (\hat{\alpha}_n, \hat{\beta}_n)$, we can also bound $M_n(\theta_0)$ using a similar argument as in the proof of Proposition 3.1 to obtain

$$
|\hat{\gamma}_n - \gamma_0|^2 \mathbb{P}_n(Z < a) \wedge \mathbb{P}_n(Z > b)
$$
$$
\leq \mathbb{P}_n\left(\epsilon^2 \mathbf{1}_{a \leq Z \leq b}\right) + 2|\hat{\gamma}_n - \gamma_0|\left(|\mathbb{P}_n\left(\epsilon \mathbf{1}_{Z < a}\right)| + |\mathbb{P}_n\left(\epsilon \mathbf{1}_{Z > b}\right)|\right).
$$

By the strong law of large numbers

$$\mathbb{P}_n(Z < a) \wedge \mathbb{P}_n(Z > b) \xrightarrow{a.s.} \mathbb{P}(Z < a) \wedge \mathbb{P}(Z > b)$$

$$\mathbb{P}_n\left(\epsilon^2 \mathbf{1}_{a \leq Z \leq b}\right) \xrightarrow{a.s.} \sigma^2 \mathbb{P}\left(a \leq Z \leq b\right) \quad \text{and}$$

$$|\mathbb{P}_n\left(\epsilon \mathbf{1}_{Z < a}\right)| + |\mathbb{P}_n\left(\epsilon \mathbf{1}_{Z > b}\right)| \xrightarrow{a.s.} 0.$$

Therefore, w.p. 1 we can write

$$|\hat{\gamma}_n - \gamma_0|^2 \leq O(1) + |\hat{\gamma}_n - \gamma_0| o(1)$$

and thus the sequence $(\hat{\gamma}_n - \gamma_0)_{n=1}^{\infty}$ is bounded w.p. 1.

Now, take any compact set $K \subset \Theta$ and consider the classes of functions

$$
\begin{aligned}
\mathcal{K}_1 &= \left\{(\epsilon + \alpha_0 - \alpha)^2 \mathbf{1}_{(-\infty, \zeta \wedge \zeta_0]}\right\}_{\theta \in K} \\
\mathcal{K}_2 &= \left\{(\epsilon + \beta_0 - \alpha)^2 \mathbf{1}_{(\zeta_0, \zeta]}\right\}_{\theta \in K} \\
\mathcal{K}_3 &= \left\{(\epsilon + \alpha_0 - \beta)^2 \mathbf{1}_{(\zeta, \zeta_0]}\right\}_{\theta \in K} \\
\mathcal{K}_4 &= \left\{(\epsilon + \beta_0 - \beta)^2 \mathbf{1}_{(\zeta \vee \zeta_0, \infty)}\right\}_{\theta \in K}.
\end{aligned}
$$

If $A^*$ is an upper bound for the norm of the elements in $K$, we can see that each of these classes is a VC-subgraph class with integrable envelope $(|\epsilon| + A^* + |\gamma_0|)^2$. With the notation $\|Q\|_{\mathcal{F}} = \sup\{|Qf| : f \in \mathcal{F}\}$ for classes of functions $\mathcal{F}$ and probability measures $Q$, a combination of Theorems 2.6.7 and 2.4.3 of Van der Vaart and Wellner (1996) shows that all four quantities $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{K}_j}$, $j = 1, 2, 3, 4$, converge to zero almost surely. Therefore using (54), we get the inequality

$$\|M_n - M\|_K \leq \sum_{1 \leq j \leq 4} \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{K}_j}$$

which now implies $(i)$ ( Since $M_n, M \in \mathcal{D}_K$, $\|M_n - M\|_K$ is measurable.). The second assertion follows immediately from $(ii)$.

Consider a family of compact rectangles $\Theta_n \subset \Theta_{n+1}$ such that $\Theta = \cup_{n=1}^{\infty} \Theta_n$. Then, since the sequence $(\hat{\gamma}_n - \gamma_0)_{n=1}^{\infty}$ is almost surely bounded, w.p. 1 we have that there is some $m \in \mathbb{N}$ such that $\Theta_m$ contains both $\theta_0$ and the entire sequence $(\hat{\theta}_n)_{n=1}^{\infty}$. Finally, from (31) with $\theta_n$ replaced by $\theta_0$ it is seen that

$$
\begin{aligned}
M(\theta) = -\sigma^2 &- (\alpha_0 - \alpha)^2 \mathbb{P}(Z \leq \zeta \wedge \zeta_0) - (\alpha_0 - \beta)^2 \mathbb{P}(\zeta < Z \leq \zeta_0) \\
&- (\alpha - \beta_0)^2 \mathbb{P}(\zeta_0 < Z \leq \zeta) - (\beta_0 - \beta)^2 \mathbb{P}(Z > \zeta \vee \zeta_0).
\end{aligned}
$$

As $\alpha_0 \neq \beta_0$ and $Z$ has a strictly positive density on $[a, b]$, the last equation shows that $M$ satisfies the conditions of Lemma A.2. Since the event that $M_n \to M$

46

in $\mathcal{D}_{\Theta_k}$ for all $k \in \mathbb{N}$ has probability one, Lemma A.2 allows us to conclude that $\operatorname{sargmax}(M_n) = \hat{\theta}_n \xrightarrow{a.s.} \theta_0$. $\qquad\square$

### A.2.9 Proof of Lemma 4.2

Let $\rho, \delta > 0$. We know from Corollary 3.1 that the sequences $(\sqrt{n}(\hat{\alpha}_n - \alpha_0))_{n=1}^{\infty}$, $\left(n(\hat{\zeta}_n - \zeta_0)\right)_{n=1}^{\infty}$ and $\left(n\mathbb{P}_n\left(\zeta_0 - \frac{h}{n} < Z \le \zeta_0 + \frac{h}{n}\right)\right)_{n=1}^{\infty}$, for any $h > 0$, are all stochastically bounded. Thus, since $m_n = O(n)$ there is $L > 0$ such that $\mathbf{P}\left(m_n|\hat{\zeta}_n - \zeta_0| > L\right) < \rho$ and $\mathbf{P}\left(\sqrt{m_n}|\hat{\alpha}_n - \alpha_0| > L\right) < \rho$ for any $n \in \mathbb{N}$. Therefore,

$$
\mathbf{P}\left(m_n^{\gamma}\left\|\mathbb{P}_n(\hat{\zeta}_n + \frac{(\cdot)}{m_n} < Z \le \hat{\zeta}_n)\right\|_K > \delta\right)
$$
$$
\le \quad \frac{m_n^{\gamma}}{\delta}\mathbf{E}\left(\mathbb{P}_n\left(\zeta_0 - \frac{L+\eta}{m_n} < Z \le \zeta_0 + \frac{L}{m_n}\right)\right) + \mathbf{P}\left(m_n|\hat{\zeta}_n - \zeta_0| > L\right)
$$
$$
\le \quad f(\zeta_0)\frac{\eta + 2L}{\delta}m_n^{\gamma-1} + o\left(m_n^{\gamma-1}\right) + \rho,
$$

so by letting $n \to \infty$ and then $\rho \to 0$ we get $(i)$.
We prove $(ii)$ for when $p = 1$, the case $p = 2$ follows from similar arguments. Note that if $m_n|\hat{\zeta}_n - \zeta_0| \le L$, then $m_n^{\gamma}\|\mathbb{P}_n(|\tilde{\epsilon}_n|\mathbf{1}_{\hat{\zeta}_n + \frac{(\cdot)}{m_n} < Z \le \hat{\zeta}_n})\|_K$ can be bounded by

$$
m_n^{\gamma}\left\|\mathbb{P}_n\left(|\epsilon|\mathbf{1}_{\zeta_0 - \frac{L}{m_n} + \frac{(\cdot)}{m_n} < Z \le \zeta_0 + \frac{L}{m_n}}\right)\right\|_K + m_n^{\gamma}|\hat{\alpha}_n - \alpha_0|\left\|\mathbb{P}_n(\zeta_0 - \frac{L}{m_n} + \frac{(\cdot)}{m_n} < Z \le \zeta_0 + \frac{L}{m_n})\right\|_K.
$$

But just as in the proof of $(i)$, we have

$$
\mathbf{P}\left(m_n^{\gamma}\left\|\mathbb{P}_n(|\epsilon|\mathbf{1}_{\hat{\zeta}_n + \frac{(\cdot)}{m_n} < Z \le \hat{\zeta}_n})\right\|_K > \delta\right)
$$
$$
\le \quad \mathbf{P}\left(m_n^{\gamma}\left\|\mathbb{P}_n\left(|\epsilon|\mathbf{1}_{\zeta_0 - \frac{L}{m_n} + \frac{(\cdot)}{m_n} < Z \le \zeta_0 + \frac{L}{m_n}}\right)\right\|_K > \frac{\delta}{2}\right) +
$$
$$
\quad \mathbf{P}\left(m_n^{\gamma}|\hat{\alpha}_n - \alpha_0|\mathbb{P}_n(\zeta_0 - \frac{L}{m_n} + \frac{\eta}{m_n} < Z \le \zeta_0 + \frac{L}{m_n}) > \frac{\delta}{2}\right) + \mathbf{P}\left(m_n|\hat{\zeta}_n - \zeta_0| > L\right)
$$
$$
\le \quad \frac{2m_n^{\gamma}}{\delta}\mathbf{E}\left(\mathbb{P}_n\left(|\epsilon|\mathbf{1}_{\zeta_0 - \frac{L}{m_n} + \frac{\eta}{m_n} < Z \le \zeta_0 + \frac{L}{m_n}}\right)\right) +
$$
$$
\quad \mathbf{P}\left(m_n^{\gamma}|\hat{\alpha}_n - \alpha_0|\mathbb{P}_n(\zeta_0 - \frac{L}{m_n} + \frac{\eta}{m_n} < Z \le \zeta_0 + \frac{L}{m_n}) > \frac{\delta}{2}\right) + \mathbf{P}\left(m_n|\hat{\zeta}_n - \zeta_0| > L\right)
$$
$$
\le \quad f(\zeta_0)\mathbf{E}\left(|\epsilon|\right)\frac{2(\eta + 2L)}{\delta}m_n^{\gamma-1} + o\left(m_n^{\gamma-1}\right) +
$$
$$
\quad \mathbf{P}\left(m_n^{\gamma}|\hat{\alpha}_n - \alpha_0|\mathbb{P}_n(\zeta_0 - \frac{L}{m_n} + \frac{\eta}{m_n} < Z \le \zeta_0 + \frac{L}{m_n}) > \frac{\delta}{2}\right) + \rho.
$$

The result follows again by letting $n \to \infty$ and $\rho \to 0$. $\qquad\square$

The next results will be useful to support our conjecture of inconsistency of some of our bootstrap scenarios.

**Lemma A.7** *Let $\lambda, B > 0$, $\rho \in (0, \frac{1}{2})$ and $H_\lambda$ be the distribution function of a Poisson random variable with mean $\lambda$. For each value of $\lambda$ write $L^\rho_{\lambda+B} = \min\{n \in \mathbb{N} : H_{\lambda+B}(n) > \rho\}$ and $U^\rho_\lambda = \max\{n \in \mathbb{N} : 1 - H_\lambda(n) > \rho\}$. Then, there is $\lambda_* > 0$ such that $L^\rho_{\lambda+B} < U^\rho_\lambda$ for all $\lambda \geq \lambda_*$.*

**Proof:** Let $c_\lambda$ be the median (i.e. $c_\lambda = \min\{n \in \mathbb{N} : H_\lambda(n) > \frac{1}{2}\}$.) of $H_\lambda$. Observe that $c_\lambda \leq U^\rho_\lambda$. According to Hazma (1995), $|c_\lambda - \lambda| < \log(2)$ for any positive $\lambda$. Letting $\lfloor x \rfloor$ denote the greatest integer less than or equal to $x$, we have

$$
\begin{aligned}
&|H_{\lambda+B}(c_{\lambda+B}) - H_{\lambda+B}(c_\lambda)| \\
\leq\ & |H_{\lambda+B}(\lambda + B + \log(2)) - H_{\lambda+B}(\lambda - \log(2))| \\
\leq\ & (B + 2\log(2))e^{-(\lambda+B)}\frac{(\lambda + B)^{\lfloor \lambda+B \rfloor}}{\lfloor \lambda + B \rfloor!} \to 0 \text{ as } \lambda \to \infty.
\end{aligned}
$$

as the Poisson mass function has a maximum at $\lfloor \lambda + B \rfloor$. Therefore, $\varliminf_{\lambda\to\infty} H_{\lambda+B}(U^\rho_\lambda) \geq 1/2$. But we also note that $\sup_{n\in\mathbb{N}}\{H_{\lambda+B}(n+1) - H_{\lambda+B}(n)\} \to 0$ as $\lambda \to \infty$. Thus,

$$
\varlimsup_{\lambda\to\infty} H_{\lambda+B}(L^\rho_{\lambda+B} + 1) = \rho < \frac{1}{2} \leq \varliminf_{\lambda\to\infty} H_{\lambda+B}(U^\rho_\lambda).
$$

It follows that $U^\rho_\lambda > L^\rho_{\lambda+B}$ for all $\lambda$ sufficiently large. $\qquad\square$

**Lemma A.8**

*Let $\lambda, B > 0$, $0 < \rho < \frac{1}{2}$, $\mu$ and $\nu$ be two nondegenerate Borel probability measures on $\mathbb{R}$ and $H_{\mu,\lambda}$ denote the compound Poisson distribution with intensity $\lambda$ and compounding distribution $\mu$. For each value of $\lambda$ write $L^\rho_{\nu,\lambda+B} = \inf\{s \in \mathbb{R} : H_{\nu,\lambda+B}(s) \geq \rho\}$ and $U^\rho_{\mu,\lambda} = \sup\{s \in \mathbb{R} : 1 - H_{\mu,\lambda}(s) \geq \rho\}$. In addition, assume that $\int x^2\nu(dx), \int x^2\mu(dx) < \infty$ and that $\int x\nu(dx) \leq \int x\mu(dx)$. Then there is $\lambda_* > 0$ such that $L^\rho_{\nu,\lambda+B} < U^\rho_{\mu,\lambda}$ for all $\lambda \geq \lambda_*$. Moreover, let $0 < r < 1$, suppose that there is another Borel probability measure $\gamma$ on $\mathbb{R}$ and define $\nu_\gamma := \frac{rB}{\lambda+B}\gamma + \frac{\lambda+(1-r)B}{\lambda+B}$ and the corresponding constant $L^\rho_{\nu_\gamma,\lambda+B} = \inf\{s \in \mathbb{R} : H_{\nu_\gamma,\lambda+B}(s) \geq \rho\}$. Then there is $\lambda_* > 0$ such that $L^\rho_{\nu_\gamma,\lambda+B} < U^\rho_{\mu,\lambda}$ for all $\lambda \geq \lambda_*$.*

**Proof:** Denote by $\Phi$ the standard normal distribution and $\mathbf{z}_\alpha$ the lower $\alpha$-quantile of $\Phi$ (i.e. $\Phi(\mathbf{z}_\alpha) = \alpha$). Also, write $c_\mu := \int x\mu(dx)$, $d_\mu := \int x^2\mu(dx)$ and define the corresponding quantities $c_\nu$ and $d_\nu$ for $\nu$. For any possible value of $\lambda$ and $\mu$ denote by $T_{\mu,\lambda}$ a random variable with distribution $H_{\mu,\lambda}$. It is easily seen (as, for instance,

in Theorem 2.1 of Möhle (2005)) that $S_{\mu,\lambda} := \frac{T_{\mu,\lambda} - \lambda c_\mu}{\sqrt{\lambda d_\mu}} \rightsquigarrow \Phi$ as $\lambda \to \infty$. Since the standard normal distribution is continuous, the distributions of $S_{\mu,\lambda}$ converge uniformly on $\mathbb{R}$ to $\Phi$ as $\lambda \to \infty$.

Let $1 < \kappa < 1/(2\rho)$. Then, since the distributions of $S_{\mu,\lambda}$ converge uniformly to $\Phi$, there is $\lambda_1$ such that $1 - \Phi\left(\frac{U^\rho_{\mu,\lambda} - \lambda c_\mu}{\sqrt{\lambda d_\mu}}\right) < \kappa\rho$ for $\lambda > \lambda_1$ and $\lambda_2 > 0$ such that $\Phi\left(\frac{L^\rho_{\nu,\lambda+B} - (\lambda+B)c_\nu}{\sqrt{(\lambda+B)d_\nu}}\right) < \kappa\rho$ for all $\lambda > \lambda_2$. These two inequalities in turn imply that

$$U^\rho_{\mu,\lambda} > \lambda c_\mu - \sqrt{\lambda d_\mu}\mathbf{z}_{\kappa\rho},$$
$$L^\rho_{\nu,\lambda+B} < (\lambda + B)c_\nu + \sqrt{(\lambda + B)d_\nu}\mathbf{z}_{\kappa\rho}.$$

Since $c_\mu \geq c_\nu$ we can find $\lambda_3$ such that

$$(\lambda + B)c_\nu + \sqrt{(\lambda + B)d_\nu}\mathbf{z}_{\kappa\rho} < \lambda c_\mu - \sqrt{\lambda d_\mu}\mathbf{z}_{\kappa\rho} \quad \text{for all } \lambda \geq \lambda_3.$$

The first part of the result now follows by taking $\lambda_* := \lambda_1 \vee \lambda_2 \vee \lambda_3$. To prove the result for the measure $\nu_\gamma$ it suffices to see that we also have $\frac{T_{\nu_\gamma,\lambda+B} - (\lambda + B)c_{\nu_\gamma}}{\sqrt{(\lambda + B)d_{\nu_\gamma}}} \rightsquigarrow \Phi$, as $\lambda \to \infty$ (this is easily seen by analyzing the characteristic functions). The rest follows from the same argument used to prove the first part of the lemma. $\square$

### A.2.10   Proof of Lemma 4.4

**Proof of** $(i)$**:** Let $s < t$. Note that $(Z_n)_{n=1}^\infty$ is a collection of i.i.d. random variables and $n\mathbb{P}_n(\zeta_0 + \frac{s}{n} < Z \leq \zeta_0 + \frac{t}{n})$ is permutation invariant, so the Hewitt-Savage 0-1 law (see page 304 of Billingsley (1986)) implies that any convergent subsequence must converge to a constant. On the other hand, Lemma 3.3 implies that $n\mathbb{P}_n(\zeta_0 + \frac{s}{n} < Z \leq \zeta_0 + \frac{t}{n}) \rightsquigarrow \text{Poisson}((t - s)f(\zeta_0))$. Therefore, $\left(n\mathbb{P}_n(\zeta_0 + \frac{s}{n} < Z \leq \zeta_0 + \frac{t}{n})\right)_{n=1}^\infty$ has no almost surely convergent subsequence.

**Proof of** $(ii)$**:** Now, let $\delta \in (0, \frac{1}{4})$. From Proposition 3.2 we know that there is $B_\delta > 0$ such that $\mathbf{P}\left(n|\hat{\zeta}_n - \zeta_0| \leq B_\delta\right) > 1 - \delta$ for any $n \in \mathbb{N}$. Choose $h > 2B_\delta$ and take any increasing sequence of natural numbers $n_k$. Write $\hat{T}_k = n_k\mathbb{P}_{n_k}(\hat{\zeta}_{n_k} < Z \leq \hat{\zeta}_{n_k} + \frac{h}{n_k})$, $S_k = n_k\mathbb{P}_{n_k}(\zeta_0 - \frac{B_\delta}{n_k} < Z \leq \zeta_0 + \frac{h+B_\delta}{n_k})$ and $T_k = n_k\mathbb{P}_{n_k}(\zeta_0 + \frac{B_\delta}{n_k} < Z \leq \zeta_0 + \frac{h-B_\delta}{n_k})$. Then, $\left\{n_k|\hat{\zeta}_{n_k} - \zeta_0| \leq B_\delta\right\} \subset \left\{S_k \geq \hat{T}_k \geq T_k\right\}$ and therefore we have $\mathbf{P}\left(\hat{T}_k \geq T_k\right) \wedge \mathbf{P}\left(S_k \geq \hat{T}_k\right) > 1 - \delta$ for all $k$.

We know that $T_k \rightsquigarrow \text{Poisson}((h - 2B_\delta)f(\zeta_0))$ and $S_k \rightsquigarrow \text{Poisson}((h + 2B_\delta)f(\zeta_0))$, so in view of Lemma A.7 with $B = 4B_\delta f(\zeta_0)$ and $\lambda = (h - 2B_\delta)f(\zeta_0)$, there

is a number $h_* > 2B_\delta$ large enough so that whenever $h \geq h_*$ we can find two numbers $N_{1,h} < N_{2,h} \in \mathbb{N}$ with the property that, $\underline{\lim}_{k\to\infty} \mathbf{P}(T_k > N_{2,h}) > 2\delta$ and $\underline{\lim}_{k\to\infty} \mathbf{P}(S_k \leq N_{1,h}) > 2\delta$. Thus, for $h \geq h_*$, $\mathbf{P}(T_k > N_{2,h}) > 2\delta$ and $\mathbf{P}(S_k \leq N_{1,h}) > 2\delta$ for all but a finite number of $k$'s. Therefore, for any $k$ large enough, $\mathbf{P}(T_k > N_{2,h}) \wedge \mathbf{P}(S_k \leq N_{1,h}) > 2\delta$. Using the fact that $\mathbf{P}\left(S_k \geq \hat{T}_k \geq T_k\right) > 1 - \delta$ we get that $\mathbf{P}\left(\hat{T}_k \geq T_k > N_{2,h}\right) \wedge \mathbf{P}\left(N_{1,h} \geq S_k \geq \hat{T}_k\right) > \delta$ for all but finitely many $k$'s. Thus, whenever $h \geq h_*$,

$$\mathbf{P}\left(\hat{T}_k \geq T_k > N_{2,h}, i.o.\right) > \delta \text{ and } \mathbf{P}\left(N_{1,h} \geq S_k \geq \hat{T}_k, i.o.\right) > \delta.$$

But for every $k \in \mathbb{N}$, the events $\left\{\hat{T}_k \geq T_k > N_{2,h}\right\}$ and $\left\{N_{1,h} \geq S_k \geq \hat{T}_k\right\}$ are permutation-invariant on the i.i.d. random vectors $X_1, \ldots, X_{n_k}$. Hence, the Hewitt-Savage 0-1 law implies that $\mathbf{P}\left(\hat{T}_k \geq T_k > N_{2,h}, i.o.\right) = 1$ and $\mathbf{P}\left(N_{1,h} \geq S_k \geq \hat{T}_k, i.o.\right) = 1$. Since $N_{1,h} < N_{2,h}$ it follows that $\hat{T}_k = n_k \mathbb{P}_{n_k}(\hat{\zeta}_{n_k} < Z \leq \hat{\zeta}_{n_k} + h/n_k)$ does not have an almost sure limit. But the choice of the subsequence $n_k$ was arbitrary and independent of $h_*$ so we can conclude that for any $h \geq h_*$, the sequence $\left\{n\mathbb{P}_n(\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{h}{n})\right\}_{n=1}^{\infty}$ does not converge in probability. Proceeding analogously, we can prove the same for $\left\{n\mathbb{P}_n(\hat{\zeta}_n - \frac{h}{n} < Z \leq \hat{\zeta}_n)\right\}_{n=1}^{\infty}$.

**Proof of** $(iii)$: We introduce some notation, for any two Borel probability measures $\mu$ and $\nu$ on $\mathbb{R}$ we write $\mu \bigstar \nu$ for their convolution and for $\lambda > 0$ we write $\text{CPoisson}(\mu, \lambda)$ for the compound Poisson distribution with intensity $\lambda$ and compounding distribution $\mu$. Let $\mu_\alpha$ and $\mu_\beta$ be, respectively, the distributions under $\mathbb{P}$ of $\phi(\epsilon + \alpha_0)$ and $\phi(\epsilon + \beta_0)$.

Observe that depending on whether $t < 0$, $s < 0 < t$ or $s > 0$ we have that $n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\zeta_0 + \frac{s}{n} < Z \leq \zeta_0 + \frac{t}{n}})$ converges weakly to $\text{CPoisson}(\mu_\alpha, (t-s)f(\zeta_0))$, $\text{CPoisson}(\mu_\alpha, sf(\zeta_0)) \bigstar \text{CPoisson}(\mu_\beta, tf(\zeta_0))$ or $\text{CPoisson}(\mu_\beta, (t-s)f(\zeta_0))$, respectively. This follows easily from convergence of the corresponding characteristic functions. Considering that $\{(Y_n, Z_n)\}_{n=1}^{\infty}$ is a collection of i.i.d. random vectors and that $n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\zeta_0 + \frac{s}{n} < Z \leq \zeta_0 + \frac{t}{n}})$ is permutation invariant for $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ the same argument as in (i) applies here as well.

**Proof of (iv):** We keep the notation used in the proof of $(iii)$. The argument here is quite similar to the one used to show $(ii)$. Assume without loss of generality that $\phi \leq 0$.

Now, let $\delta \in \left(0, \frac{1}{4}\right)$ and $N \in \mathbb{N}$. From Proposition 3.2 we know that there is $B_\delta > 0$ such that $\mathbf{P}\left(n|\hat{\zeta}_n - \zeta_0| \leq B_\delta\right) > 1 - \delta$ for any $n \in \mathbb{N}$. Choose $h > 2B_\delta$ and take any increasing sequence of natural numbers $n_k$. Write $\hat{T}_{k,h}^\phi =$

$n_k \mathbb{P}_{n_k}(\phi(Y)\mathbf{1}_{\hat{\zeta}_{n_k} < Z \le \hat{\zeta}_{n_k} + \frac{h}{n_k}})$, $S_{k,h}^\phi = n_k \mathbb{P}_{n_k}(\phi(Y)\mathbf{1}_{\zeta_0 - \frac{B_\delta}{n_k} < Z \le \zeta_0 + \frac{h + B_\delta}{n_k}})$ and $T_{k,h}^\phi = n_k \mathbb{P}_{n_k}(\phi(Y)\mathbf{1}_{\zeta_0 + \frac{B_\delta}{n_k} < Z \le \zeta_0 + \frac{h - B_\delta}{n_k}})$. Then, $\left\{ n_k |\hat{\zeta}_{n_k} - \zeta_0| \le B_\delta \right\} \subset \left\{ S_{k,h}^\phi \le \hat{T}_{k,h}^\phi \le T_{k,h}^\phi \right\}$ and therefore we have $\mathbf{P}\left( \hat{T}_{k,h}^\phi \le T_{k,h}^\phi \right) \wedge \mathbf{P}\left( S_{k,h}^\phi \le \hat{T}_{k,h}^\phi \right) > 1 - \delta$ for all $k$.

We know that $T_{k,h}^\phi \rightsquigarrow \text{CPoisson}(\mu_\beta, (h - 2B_\delta)f(\zeta_0))$ and

$$S_k^\phi \rightsquigarrow \text{CPoisson}(\mu_\alpha, 2B_\delta f(\zeta_0)) \bigstar \text{CPoisson}(\mu_\beta, (h + B_\delta)f(\zeta_0))$$

$$\equiv \text{CPoisson}\left( \frac{B_\delta}{h + 2B_\delta}\mu_\alpha + \frac{h + B_\delta}{h + 2B_\delta}\mu_\beta, (h + 2B_\delta)f(\zeta_0) \right),$$

as $k \to \infty$.

An application of Lemma A.8 with $\mu = \nu = \mu_\beta$, $\gamma = \mu_\alpha$, $B = 4B_\delta f(\zeta_0)$, $r = \frac{1}{4}$ and $\lambda = (h - 2B_\delta)f(\zeta_0)$, shows the existence of an $h_* > 2B_\delta$ large enough so that whenever $h \ge h_*$ we can find two numbers $R_{1,h} > R_{2,h} \in \mathbb{N}$ with the property that $\varliminf_{k\to\infty} \mathbf{P}\left( T_{k,h}^\phi < R_{2,h} \right) > 2\delta$ and $\varliminf_{k\to\infty} \mathbf{P}\left( S_{k,h}^\phi \ge R_{1,h} \right) > 2\delta$. Thus, for $h \ge h_*$, $\mathbf{P}\left( T_{k,h}^\phi < R_{2,h} \right) > 2\delta$ and $\mathbf{P}\left( S_{k,h}^\phi \ge R_{1,h} \right) > 2\delta$ for all but a finite number of $k$'s. Therefore, for any $k$ large enough, $\mathbf{P}\left( T_{k,h}^\phi < R_{2,h} \right) \wedge \mathbf{P}\left( S_{k,h}^\phi \ge R_{1,h} \right) > 2\delta$. Using the fact that $\mathbf{P}\left( S_{k,h}^\phi \le \hat{T}_{k,h}^\phi \le T_{k,h}^\phi \right) > 1 - \delta$ we get that $\mathbf{P}\left( \hat{T}_{k,h}^\phi \le T_{k,h}^\phi < R_{2,h} \right) \wedge \mathbf{P}\left( R_{1,h} \le S_{k,h}^\phi \le \hat{T}_{k,h}^\phi \right) > \delta$ for all but finitely many $k$'s. Thus, whenever $h \ge h_*$,

$$\mathbf{P}\left( \hat{T}_{k,h}^\phi \le T_{k,h}^\phi < R_{2,h}, i.o. \right) > \delta \text{ and } \mathbf{P}\left( R_{1,h} \le S_{k,h}^\phi \le \hat{T}_{k,h}^\phi, i.o. \right) > \delta.$$

The argument relying on the Hewitt-Savage 0-1 law applied in the proof of $(ii)$ can be used to finish this proof.

A completely analogous proof applies for $\left\{ n\mathbb{P}_n(\phi(Y)\mathbf{1}_{\hat{\zeta}_n - \frac{h}{n} < Z \le \hat{\zeta}_n}) \right\}_{n=1}^\infty$. $\qquad\square$

### A.2.11   Proof of Lemma 4.6

We start by computing the characteristic functions of the weak limits of the last two components of the process $\tilde{\Xi}_n$ as defined in (17). Let $g_n(\xi)$ and $\psi_n(\xi)$ be the (unconditional) characteristic functions of $n\mathbb{P}_n^*(\mathbf{1}_{\zeta_0 < Z \le \zeta_0 + \frac{t}{n}})$ and $n\mathbb{P}_n^*(\epsilon\mathbf{1}_{\zeta_0 < Z \le \zeta_0 + \frac{t}{n}})$, respectively. Fix $\xi \in \mathbb{R}$ and write

$$\begin{aligned}
\Lambda_n &:= \mathbf{E}_{\mathfrak{X}}\left( e^{i\xi n\mathbb{P}_n^*(\epsilon\mathbf{1}_{\zeta_0 < Z \le \zeta_0 + \frac{t}{n}})} \right), \\
\Psi_n &:= n\mathbb{P}_n\left( \left(e^{i\xi\epsilon} - 1\right)\mathbf{1}_{\zeta_0 < Z \le \zeta_0 + \frac{t}{n}} \right), \\
\Psi_\xi^* &:= \sum_{1 \le k \le \nu(t)} \left( e^{i\xi\epsilon_k} - 1 \right),
\end{aligned}$$

where $(\nu(s))_{s \geq 0}$ is a Poisson process with rate $f(\zeta_0)$ independent of $(\epsilon_n)_{n=1}^{\infty}$. Then, $\psi_n(\xi) = \mathbf{E}(\Lambda_n)$ and $|\Lambda_n| \leq 1$. By the conditional independence of the bootstrap samples, we have

$$\Lambda_n = \left(1 + \frac{1}{n}\Psi_n\right)^n.$$

We now consider the characteristic functions of the complex-valued random variables $\Psi_n$. Taking into account the independence of the $X$'s, we obtain that for any $\eta \in \mathbb{R}^2$,

$$\mathbf{E}\left(e^{i\eta_1 \mathrm{Re}(\Psi_n)+i\eta_2 \mathrm{Im}(\Psi_n)}\right) = \left(1 + \frac{1}{n}\mathbb{P}\left(e^{i\eta_1(\cos(\xi\epsilon)-1)+i\eta_2\sin(\xi\epsilon)} - 1\right)\left(n\mathbb{P}(\mathbf{1}_{\zeta_0 < Z < \zeta_0 + \frac{t}{n}})\right)\right)^n$$

$$\mathbf{E}\left(e^{i\eta_1 \mathrm{Re}(\Psi_n)+i\eta_2 \mathrm{Im}(\Psi_n)}\right) \to e^{tf(\zeta_0)\mathbf{E}\left(e^{i\eta_1(\cos(\xi\epsilon)-1)+i\eta_2\sin(\xi\epsilon)}-1\right)} = \mathbf{E}\left(e^{i\eta_1 \mathrm{Re}\Psi_\xi^* + i\eta_2 \mathrm{Im}\Psi_\xi^*}\right).$$

Therefore, $\Psi_n \rightsquigarrow \Psi_\xi^*$ and, from the continuous mapping theorem, $\Lambda_n \rightsquigarrow e^{\Psi_\xi^*}$. Thus, Lebesgue's Dominated Convergence Theorem implies

$$\psi_n(\xi) = \mathbf{E}(\Lambda_n) \to \mathbf{E}\left(e^{\Psi_\xi^*}\right) = e^{tf(\zeta_0)\left(\mathbf{E}\left(e^{e^{i\xi\epsilon}-1}\right)-1\right)} \quad \forall\, \xi \in \mathbb{R}. \tag{55}$$

With simpler arguments, we can also show that

$$g_n(\xi) \to e^{tf(\zeta_0)\left(e^{(e^{i\xi\epsilon}-1)}-1\right)} \quad \forall\, \xi \in \mathbb{R}. \tag{56}$$

While (56) is immediately recognized as the characteristic function of a compound Poisson process with rate $f(\zeta_0)$ and compounding distribution Poisson(1), the characteristic function in (55) can be shown to correspond to another compound Poisson process which can be written as

$$\sum_{1 \leq j \leq \nu(t)} \epsilon_j \tau_j, \tag{57}$$

where $(\tau_n)_{n=1}^{\infty} \overset{i.i.d.}{\sim} \mathrm{Poisson}(1)$, $(\nu(s))_{s \geq 0}$ is a Poisson process with rate $f(\zeta_0)$, and $(\tau_n)_{n=1}^{\infty}$, $(\epsilon_n)_{n=1}^{\infty}$ and $(\nu(s))_{s \geq 0}$ are mutually independent.

Therefore, the fifth and sixth components of $\tilde{\Xi}_n$ as defined in (17) converge, respectively, to a compound Poisson process with rate $f(\zeta_0)$ and Poisson(1) as compounding distribution and to the process described in (57). A similar analysis shows the analogous results for the third and fourth components of $\tilde{\Xi}_n$. The first and second components of $\tilde{\Xi}_n$ can easily be seen (by using the Lindeberg-Feller Central Limit Theorem) to be asymptotically normal with mean 0 and variances $\sigma^2 \mathbb{P}(Z \leq \zeta_0)$ and $\sigma^2 \mathbb{P}(Z > \zeta_0)$, respectively.

All these facts indicate that the finite dimensional distributions of the limiting process of $\tilde{\Xi}_n$ match those of the process $\tilde{\tilde{\Xi}}$. In fact, we can proceed as in the proof of Proposition 3.3 (i.e., proving tightness and convergence of the finite dimensional distributions using the Cramer-Wold device) to show $(i)$ and $(ii)$. For the sake of brevity, we omit the full technical details.

Then, arguing as in Proposition 3.2 one can show that the sequence $(\sqrt{n}(\alpha_n^* - \alpha_0), \sqrt{n}(\beta_n^* - \beta_0), n(\zeta_n^* - \zeta_0))$ is stochastically bounded and then conclude that the (unconditional) asymptotic distribution of $(\sqrt{n}(\alpha_n^* - \alpha_0), \sqrt{n}(\beta_n^* - \beta_0), n(\zeta_n^* - \zeta_0)$ is that of $\mathrm{sargmax}_{h \in \mathbb{R}^3}\{\tilde{E}^*(h)\}$, with $\tilde{E}^*(h)$ as defined in (18) and (19). For the sake of brevity we omit the full technical details of these arguments.

As $n(\zeta_n^* - \zeta_0) = n(\zeta_n^* - \hat{\zeta}_n) + n(\hat{\zeta}_n - \zeta_0)$, and if the ECDF bootstrap were consistent, the conditional distribution of $n(\zeta_n^* - \hat{\zeta}_n)$ (given the data) and the unconditional distribution of $n(\hat{\zeta}_n - \zeta_0)$ would have had the same weak limit. Then, as a consequence of Lemma 3.1 in Sen et al. (2010) (also see Theorem 2.2 in Kosorok (2008a)) the unconditional asymptotic distribution of $n(\zeta_n^* - \zeta_0)$ must be that of the sum of two independent copies of the asymptotic distribution of the $n(z\hat{e}ta_n - \zeta_0)$. The result now follows. $\qquad \square$

### A.2.12  Proof of Lemma 4.7

Let $\mathbb{G}_n$ be the ECDF of $\epsilon_1, \ldots, \epsilon_n$. We first observe that

$$\int e^{i\xi x} d\mathbb{P}_n^\epsilon(x) = e^{-i\xi \bar{\epsilon}_n} \mathbb{P}_n\left(e^{i\xi \widetilde{\epsilon}_n}\right)$$

and hence, for any $\xi \in \mathbb{R}$ with $|\xi| \leq \eta$ we have,

$$\left| \int e^{i\xi x} d\mathbb{P}_n^\epsilon(x) - e^{-i\xi \bar{\epsilon}_n} \int e^{i\xi x} d\mathbb{G}_n(x) \right| = \left| \mathbb{P}_n\left(e^{i\xi \widetilde{\epsilon}_n}\right) - \mathbb{P}_n\left(e^{i\xi \epsilon}\right) \right|$$
$$\leq |\eta| \mathbb{P}_n\left(|\tilde{\epsilon}_n - \epsilon|\right)$$

but $\mathbb{P}_n\left(|\tilde{\epsilon}_n - \epsilon|\right)$ is bounded from above by

$$|\hat{\alpha}_n - \alpha_0| + (|\alpha_0| + |\beta_0|)\left|\mathbb{P}_n(\mathbf{1}_{Z \leq \hat{\zeta}_n} - \mathbf{1}_{Z \leq \zeta_0})\right| + |\hat{\beta}_n - \beta_0|$$

which goes to zero almost surely as consequence of Lemmas 4.1 and A.5 $(iv)$, with $\mathbb{Q}_n = \mathbb{P}_n$. Thus,

$$\sup_{|\xi| \leq \eta}\left\{\left| \int e^{i\xi x} d\mathbb{P}_n^\epsilon(x) - e^{-i\xi \bar{\epsilon}_n} \int e^{i\xi x} d\mathbb{G}_n(x) \right|\right\} \xrightarrow{a.s.} 0$$

and $(i)$ follows immediately because $\bar{\epsilon}_n = \mathbb{P}_n(\tilde{\epsilon}_n) \xrightarrow{a.s.} 0$ and $\mathbb{G}_n$ converges to $G$ in total variation distance with probability one. The second assertion is seen to be

true at once because $G$ is assumed to be continuous and condition $(i)$ implies that the characteristic functions of $\mathbb{P}_n^\epsilon$ converge to the characteristic function of $G$ on the entire real line with probability one. Statements $(ii)$ and $(iii)$ are straightforward: On the one hand, we have shown that conditions (I)-(IV) hold for the ECDF, so Lemma A.5 implies that $\int x^2 d\mathbb{P}_n^\epsilon(x) = \mathbb{P}_n(\tilde{\epsilon}_n^2) - \mathbb{P}_n(\tilde{\epsilon}_n)^2 \xrightarrow{a.s.} \sigma^2$. On the other hand,

$$\left| \int |x| d\mathbb{P}_n^\epsilon - \int |\epsilon| d\mathbb{P}_n \right| = |\mathbb{P}_n(|\tilde{\epsilon}_n - \bar{\epsilon}_n| - |\epsilon|)|$$

$$\leq \mathbb{P}_n(|\tilde{\epsilon}_n - \epsilon|) + |\bar{\epsilon}_n| \xrightarrow{a.s.} 0.$$

To prove $(iv)$, we first notice that

$$\int |x|^3 d\mathbb{P}_n^\epsilon(x) \leq |\bar{\epsilon}_n|^3 + 3|\bar{\epsilon}_n|^2 \mathbb{P}_n(|\tilde{\epsilon}_n|) + 3|\bar{\epsilon}_n| \mathbb{P}_n(\tilde{\epsilon}_n^2) + \mathbb{P}_n(|\tilde{\epsilon}_n|^3).$$

Then, from Lemma A.6 all but the last summand on the right-hand side converge almost surely. Hence, it suffices to show that $\overline{\lim} \, \mathbb{P}_n(|\tilde{\epsilon}_n|^3) < \infty$ w. p. 1. With this in mind, let $L_n = |\alpha_0| + |\hat{\alpha}_n| + |\beta_0| + |\hat{\beta}_n|$ and observe that

$$\mathbb{P}_n(|\tilde{\epsilon}_n|^3) \leq \mathbb{P}_n(|\epsilon|^3) + 3\mathbb{P}_n(|\epsilon|^2) L_n + 3\mathbb{P}_n(|\epsilon|) L_n^2 + L_n^3.$$

The result then is an immediate consequence of the third moment assumption on $\epsilon$, the strong law of large numbers and the almost sure convergence of the least squares estimators. $\qquad\square$

### A.2.13   Proof of Proposition 4.2

Just as in the proof of Proposition 3.1 we have

$$-\frac{1}{n} \sum_{k=1}^n (\tilde{\epsilon}_{n,j}^*)^2 = R_n(\hat{\theta}_n)$$

$$\leq R_n(\theta_n^*) \leq -\frac{1}{n} \sum_{j=1}^n (\tilde{\epsilon}_{n,j}^* + \hat{\alpha}_n - \alpha_n^*)^2 \mathbf{1}_{Z_j < a} + (\tilde{\epsilon}_{n,j}^* + \hat{\beta}_n - \beta_n^*)^2 \mathbf{1}_{Z_j > b}$$

from which we can see that

$$|\gamma_n^* - \gamma_n|^2 \mathbb{P}_n(Z < a) \wedge \mathbb{P}_n(Z > b) \leq$$

$$\frac{1}{n} \sum_{j=1}^n (\tilde{\epsilon}_{n,j}^*)^2 \mathbf{1}_{a \leq Z_j \leq b} + \frac{2}{n} |\gamma_n^* - \gamma_n| \left( \left| \sum_{j=1}^n \tilde{\epsilon}_{n,j}^* \mathbf{1}_{Z_j < a} \right| + \left| \sum_{j=1}^n \tilde{\epsilon}_{n,j}^* \mathbf{1}_{Z_j > b} \right| \right).$$

But the first of the terms on the right-hand side of the previous inequality is conditionally bounded in $\mathbb{L}_1$ (an upper bound for the conditional expectations is $\sup_{n\in\mathbb{N}}\left\{\int x^2 d\mathbb{P}_n^\epsilon(x)\right\} < \infty$). The terms $\frac{1}{n}\sum_{j=1}^n \tilde{\epsilon}_{n,j}^* \mathbf{1}_{Z_j<a}$ and $\frac{1}{n}\sum_{j=1}^n \tilde{\epsilon}_{n,j}^* \mathbf{1}_{Z_j>b}$ both have zero conditional expectation and conditional variances equal to $\frac{1}{n}\mathbb{P}_n(Z < a)\int x^2 d\mathbb{P}_n^\epsilon(x)$ and $\frac{1}{n}\mathbb{P}_n(Z > b)\int x^2 d\mathbb{P}_n^\epsilon(x)$ respectively. So we have that

$$\left|\frac{1}{n}\sum_{j=1}^n \tilde{\epsilon}_{n,j}^* \mathbf{1}_{Z_j<a}\right| + \left|\frac{1}{n}\sum_{j=1}^n \tilde{\epsilon}_{n,j}^* \mathbf{1}_{Z_j>b}\right| \xrightarrow[a.s.]{\mathbf{P}_{\mathfrak{x}}} 0.$$

Thus,

$$|\gamma_n^* - \hat{\gamma}_n| = O_{\mathbf{P}_{\mathfrak{x}}}(1) \text{ almost surely.} \tag{58}$$

Now, let $Z_{(k)}$ be the $k$-th order statistic from the sample $(Z_1,\ldots,Z_n)$ and $r_k$ a number such that $Z_{(k)} = Z_{r_k}$. For any $\zeta \in [a,b]$ define $m_\zeta = \max\{1 \le j \le n : Z_{(j)} \le \zeta \wedge \hat{\zeta}_n\}$ and observe that we have

$$\frac{1}{n}\sum_{j=1}^n \tilde{\epsilon}_{n,j}^* \mathbf{1}_{Z_j \le \zeta \wedge \hat{\zeta}_n} = \frac{1}{n}\sum_{1\le j\le m_\zeta} \tilde{\epsilon}_{n,r_j}^*, \tag{59}$$

and thus

$$\sup_{\zeta\in[a,b]}\left\{\left|\frac{1}{n}\sum_{j=1}^n \tilde{\epsilon}_{n,j}^* \mathbf{1}_{Z_j \le \zeta \wedge \hat{\zeta}_n}\right|\right\} \le \max_{1\le k\le n}\left\{\frac{1}{n}\left|\sum_{1\le j\le k} \tilde{\epsilon}_{n,r_j}^*\right|\right\}. \tag{60}$$

But the indexes $r_k$ and the order statistics are functions of $Z_1,\ldots,Z_n$ and therefore $\mathfrak{X}$-measurable. Hence, conditionally, $\sum_{1\le j\le k} \tilde{\epsilon}_{n,r_j}^* \mathbf{1}_{Z_{r_j} \le \zeta \wedge \hat{\zeta}_n}$ is a square integrable martingale with zero expectation. Hence, from Doob's submartingale inequality (see Williams (1991), Theorem 14.6, page 137) we get

$$\mathbf{P}_{\mathfrak{x}}\left(\max_{1\le k\le n}\left\{\frac{1}{n}\left|\sum_{1\le j\le k} \tilde{\epsilon}_{n,r_j}^*\right|\right\} > \rho\right) \le \frac{1}{n\rho^2}\mathbb{P}_n(\tilde{\epsilon}_n^2)$$

and consequently, equations (59) and (60) show that

$$\mathbf{P}_{\mathfrak{x}}\left(\left\|\frac{1}{n}\sum_{j=1}^n \tilde{\epsilon}_{n,j}^* \mathbf{1}_{Z_j \le (\cdot) \wedge \hat{\zeta}_n}\right\|_{[a,b]} > \rho\right) \le \frac{1}{\rho^2 n}\mathbb{P}_n(\tilde{\epsilon}_n^2) \xrightarrow{a.s.} 0. \tag{61}$$

Similar arguments give that (61) is also true if we replace $\mathbf{1}_{Z_j \leq (\cdot) \wedge \hat\zeta_n}$ by any of $\mathbf{1}_{(\cdot) < Z_j \leq \hat\zeta_n}$, $\mathbf{1}_{\hat\zeta_n < Z_j \leq (\cdot)}$ or $\mathbf{1}_{Z_j > (\cdot) \vee \hat\zeta_n}$. Now, if we write $R_n$ like

$$
\begin{aligned}
R_n(\theta) = {} & -\mathbb{P}_n^*(\tilde\epsilon_n^2) - \frac{2}{n}(\hat\alpha_n - \alpha) \sum_{j=1}^n \tilde\epsilon_{n,j}^* \mathbf{1}_{Z_j \leq \zeta \wedge \hat\zeta_n} - (\hat\alpha_n - \alpha)^2 \mathbb{P}_n(\mathbf{1}_{Z \leq \zeta \wedge \hat\zeta_n}) \\
& -\frac{2}{n}(\hat\beta_n - \alpha) \sum_{j=1}^n \tilde\epsilon_{n,j}^* \mathbf{1}_{\hat\zeta_n < Z \leq \zeta} - (\hat\beta_n - \alpha)^2 \mathbb{P}_n(\mathbf{1}_{\hat\zeta_n < Z \leq \zeta}) \\
& -\frac{2}{n}(\hat\alpha_n - \beta) \sum_{j=1}^n \tilde\epsilon_{n,j}^* \mathbf{1}_{\zeta < Z \leq \hat\zeta_n} - (\hat\alpha_n - \beta)^2 \mathbb{P}_n(\mathbf{1}_{\zeta < Z \leq \hat\zeta_n}) \\
& -\frac{2}{n}(\hat\beta_n - \beta) \sum_{j=1}^n \tilde\epsilon_{n,j}^* \mathbf{1}_{Z > \zeta \vee \hat\zeta_n} - (\hat\beta_n - \beta)^2 \mathbb{P}_n(\mathbf{1}_{Z > \zeta \vee \hat\zeta_n}), \quad (62)
\end{aligned}
$$

($ii$) follows immediately from (61), applied for all the four possible types of indicator functions. Note that the four terms on the far right of all the rows in the previous display vanish when we subtract $M_n$ from $R_n$. Lemma 4.1 shows that ($ii$) implies ($i$), while Corollary 3.2.3 ($ii$), page 287, of Van der Vaart and Wellner (1996) together with (58) allows one to derive ($iii$) from ($i$) and ($ii$). □

### A.2.14   Proof of Lemma 4.8

The proof is analogous to the proof of Lemma 4.5. We again consider the number $h_* > 0$ defined in the statement of Lemma 4.4 and take $K \subset \mathbb{R}^3$ to be any compact rectangle containing the point $(0, 0, h_*)$. To prove the theorem it suffices to show that the sequence $(\hat E_n(0, 0, h_3))_{n=1}^\infty$ does not have a weak limit in probability whenever $h_3 \geq h_*$ and $(0, 0, h_3) \in K$. But in view of Lemma 4.4 this is straightforward because the (conditional) characteristic function of $\hat E_n(0, 0, h_3)$ is given by

$$
\left( \int e^{i2(\hat\alpha_n - \hat\beta_n)\xi x - i\xi(\hat\alpha_n - \hat\beta_n)^2} d\mathbb{P}_n^\epsilon(x) \right)^{n\mathbb{P}_n(\hat\zeta_n < Z \leq \hat\zeta_n + \frac{h_3}{n})}.
$$

and Lemma 4.7 and the strong consistency of the least squares estimator imply that

$$
\int e^{i2(\hat\alpha_n - \hat\beta_n)\xi x - i\xi(\hat\alpha_n - \hat\beta_n)^2} d\mathbb{P}_n^\epsilon(x) \xrightarrow{a.s.} e^{-i\xi(\alpha_0 - \beta_0)^2} \varphi\left(2(\alpha_0 - \beta_0)\xi\right).
$$

Thus, for all $\xi$ in a neighborhood of the origin, this characteristic function will converge if and only if $n\mathbb{P}_n(\hat\zeta_n < Z \leq \hat\zeta_n + \frac{h_3}{n})$ converges. We know that this is not the case from Lemma 4.4. □

### A.2.15  Proof of Proposition 5.1

We will show that conditions (I)-(V) in Section 3 hold w.p. 1 for the bootstrap measures arising in this scheme. Note that (IV) is a consequence of Lemma 4.1. That $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \xrightarrow{a.s.} 0$ follows immediately from the fact that $\|\hat{F}_n - F\|_\infty \xrightarrow{a.s.} 0$. Now, for any $g = y\psi \in \mathcal{G}$ with $\psi \in \mathcal{F}$, we have

$$
\begin{aligned}
\mathbb{Q}_n(g) &= \hat{\alpha}_n \mathbb{Q}_n(\mathbf{1}_{Z \leq \hat{\zeta}_n} \psi) + \hat{\beta}_n \mathbb{Q}_n(\mathbf{1}_{Z > \hat{\zeta}_n} \psi), \\
\mathbb{P}(g) &= \alpha_0 \mathbb{P}(\mathbf{1}_{Z \leq \zeta_0} \psi) + \beta_0 \mathbb{P}(\mathbf{1}_{Z > \zeta_0} \psi),
\end{aligned}
$$

from which we see that

$$
\begin{aligned}
\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{G}} &\leq \left( |\hat{\alpha}_n - \alpha_0| + \left| \hat{\beta}_n - \beta_0 \right| \right) + (|\alpha_0| + |\beta_0|) \|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \\
&\quad + (|\alpha_0| + |\beta_0|) \int_{\mathbb{R}} |\mathbf{1}_{z \leq \hat{\zeta}_n} - \mathbf{1}_{z \leq \zeta_0}| \hat{f}_n(z) dz.
\end{aligned}
$$

Lebesgue's dominated convergence theorem shows that the last integral goes almost surely to zero and the strong consistency of the least squares estimators and property (I) now yields $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{G}} \xrightarrow{a.s.} 0$. Finally, we can write any $h \in \mathcal{H}$ in the form $h = y^2\psi$ for some $\psi \in \mathcal{F}$. Using this representation we obtain,

$$
\begin{aligned}
\mathbb{Q}_n(h) &= \hat{\alpha}_n^2 \mathbb{Q}_n(\mathbf{1}_{Z \leq \hat{\zeta}_n} \psi) + \hat{\beta}_n^2 \mathbb{Q}_n(\mathbf{1}_{Z > \hat{\zeta}_n} \psi) + \mathbb{P}_n^\epsilon(\tilde{\epsilon}_n^2) \mathbb{Q}_n(\psi), \\
\mathbb{P}(h) &= \alpha_0^2 \mathbb{P}(\mathbf{1}_{Z \leq \zeta_0} \psi) + \beta_0^2 \mathbb{P}(\mathbf{1}_{Z > \zeta_0} \psi) + \sigma^2 \mathbb{P}(\psi),
\end{aligned}
$$

and the triangle inequality then implies that

$$
\begin{aligned}
\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{H}} &\leq (|\hat{\alpha}_n^2 - \alpha_0^2| + |\hat{\beta}_n^2 - \beta_0^2|) + (\alpha_0^2 + \beta_0^2 + \sigma^2) \|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \\
&\quad + |\mathbb{P}_n^\epsilon(\tilde{\epsilon}_n^2) - \mathbb{P}(\epsilon^2)| + (\alpha_0^2 + \beta_0^2) \int_{\mathbb{R}} |\mathbf{1}_{z \leq \hat{\zeta}_n} - \mathbf{1}_{z \leq \zeta_0}| \hat{f}_n(z) dz \xrightarrow{a.s.} 0.
\end{aligned}
$$

It remains to show (V). Observe that (6) and (7) hold automatically because under $\mathbb{Q}_n$, $\tilde{\epsilon}_n$ and $Z$ are independent. Hence, we only require to show that (5) holds w.p. 1. As (23) holds, we have

$$
\inf_{\zeta \in [c,d]} \left\{ \hat{f}_n(\zeta) \right\} \xrightarrow{a.s.} \inf_{\zeta \in [c,d]} \{ f(\zeta) \} > 0.
$$

The mean value theorem implies that for any $\zeta, \xi \in [c,d]$, there is $\vartheta \in [0,1]$ such that $|\hat{F}_n(\zeta) - \hat{F}_n(\xi)| = |\xi - \zeta| \hat{f}_n(\zeta + \vartheta(\xi - \zeta))$. It follows that for $\eta > 0$ small enough,

$$
\inf_{0 < |\zeta - \hat{\zeta}_n| < \delta^2} \left\{ \frac{1}{|\zeta - \hat{\zeta}_n|} |\hat{F}_n(\zeta) - \hat{F}_n(\hat{\zeta}_n)| \right\} \geq \inf_{\zeta \in [c,d]} \left\{ \hat{f}_n(\zeta) \right\} \quad \forall\, n \in \mathbb{N}
$$

and consequently (V) holds w.p.1 for all $\delta < \eta$ for all large $n$. $\qquad\square$

### A.2.16  Proof of Proposition 5.2

We already know that conditions (I)-(V) hold w.p. 1. Condition (VII) holds automatically because $Z$ and $\tilde{\epsilon}_n$ are independent under $\mathbb{Q}_n$ and $\mathbb{Q}_n(\tilde{\epsilon}_n) = 0$. Lemma 4.7 $(v)$ implies that condition (VIII) holds a.s. It remains to prove (VI).

Write $I = [c, d]$ and consider the sequence of events $\{A_N\}_{N \in \mathbb{N}}$ given by

$$A_N = \left[\hat{\zeta}_n - \frac{\delta}{n}, \hat{\zeta}_n + \frac{\eta}{n} \in I, \text{ almost always, } \forall\, \delta, \eta \in (0, N)\right] \cap \left[\|\hat{f}_n - f\|_I \to 0\right].$$

Fix $N \in \mathbb{N}$, let $\psi$ be the function $\psi(x) = e^{i\xi x}$ for some $\xi \in \mathbb{R}$ or the function $\psi(x) = |x|^p$, $p = 1, 2$, and $\eta, \delta > 0$ be any positive real numbers smaller than $N$. Then,

$$m_n \mathbb{Q}_n(\psi(\tilde{\epsilon}_n)\mathbf{1}_{\zeta_n - \frac{\delta}{n} < Z \leq \zeta_n + \frac{\eta}{n}}) = n\mathbb{P}_n^\epsilon(\psi) \int_{\hat{\zeta}_n - \frac{\delta}{n}}^{\hat{\zeta}_n + \frac{\eta}{n}} \hat{f}_n(x)dx.$$

Lemma 4.7 implies that $\mathbb{P}_n^\epsilon(\psi) \xrightarrow{a.s.} \mathbb{P}(\psi(\epsilon))$. And, when $A_N$ holds, we also have

$$n\left|\int_{\hat{\zeta}_n - \frac{\delta}{n}}^{\hat{\zeta}_n + \frac{\eta}{n}} \hat{f}_n(x)dx - \int_{\hat{\zeta}_n - \frac{\delta}{n}}^{\hat{\zeta}_n + \frac{\eta}{n}} f(x)dx\right| \leq 2N \left\|\hat{f}_n - f\right\|_{[c,d]} \to 0.$$

Hence, condition (VI) holds for all $0 < \delta, \eta < N$ on $A_N$. But the strong consistency of the least squares estimators and the conditions on $\hat{f}_n$ imply that each of these events have probability one. Therefore, $\mathbf{P}\left(\cap_{N \in \mathbb{N}} A_N\right) = 1$. Hence, condition (VI) holds w.p.1 and the result follows from an application of Proposition 3.3.  $\square$

### A.2.17  Proof of Proposition 5.3

Since $\mathbb{Q}_n$ is just the ECDF, the validity of conditions (I)-(IV) follows from the result established for the regular ECDF bootstrap and Lemma 4.1. (VIII) is a consequence of the strong law of large numbers. It remains to show (V)-(VII).
We start with (VI). First observe that $m_n \mathbb{P}(\psi(\epsilon)\mathbf{1}_{\zeta_0 - \frac{\delta}{m_n} < Z \leq \zeta_0 + \frac{\eta}{m_n}}) \to (\delta + \eta)f(\zeta_0)\mathbb{P}(\psi(\epsilon))$.
We will proceed as follows: we will first use this simple observation just made to show that the following equations are true,

$$m_n \left\|\mathbb{P}_n(\psi(\epsilon)\mathbf{1}_{\zeta_0 - \frac{(\cdot)}{m_n} < Z \leq \zeta_0}) - (\cdot)\mathbb{P}(\psi(\epsilon))f(\zeta_0)\right\|_K \xrightarrow{\mathbf{P}} 0 \qquad (63)$$

$$m_n \left\|\mathbb{P}_n(\psi(\epsilon)\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}}) - (\cdot)\mathbb{P}(\psi(\epsilon))f(\zeta_0)\right\|_K \xrightarrow{\mathbf{P}} 0 \qquad (64)$$

$$m_n \left\|\mathbb{P}_n(\psi(\tilde{\epsilon}_n)\mathbf{1}_{\hat{\zeta}_n - \frac{(\cdot)}{m_n} < Z \leq \hat{\zeta}_n}) - \mathbb{P}_n(\psi(\epsilon)\mathbf{1}_{\zeta_0 - \frac{(\cdot)}{m_n} < Z \leq \zeta_0})\right\|_K \xrightarrow{\mathbf{P}} 0 \qquad (65)$$

$$m_n \left\|\mathbb{P}_n(\psi(\tilde{\epsilon}_n)\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}}) - \mathbb{P}_n(\psi(\epsilon)\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}})\right\|_K \xrightarrow{\mathbf{P}} 0 \qquad (66)$$

58

for any compact interval $K \subset \mathbb{R}$. All these facts put together will give

$$m_n \left\| \mathbb{P}_n(\psi(\tilde{\epsilon}_n) \mathbf{1}_{\hat{\zeta}_n - \frac{(\cdot)}{m_n} < Z \leq \hat{\zeta}_n}) - (\cdot) \mathbb{P}(\psi(\epsilon)) f(\zeta_0) \right\|_K \xrightarrow{\mathbf{P}} 0 \qquad (67)$$

$$m_n \left\| \mathbb{P}_n(\psi(\tilde{\epsilon}_n) \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}}) - (\cdot) \mathbb{P}(\psi(\epsilon)) f(\zeta_0) \right\|_K \xrightarrow{\mathbf{P}} 0 \qquad (68)$$

for any compact interval $K \subset \mathbb{R}$. Having achieved this, we will be able to conclude that (VI) holds in probability. For if (67) and (68) are both true, we can take an increasing sequence of compacts $(K_n)_{n=1}^\infty$ whose union is $\mathbb{R}$ and then for any subsequence $(n_k)_{k=1}^\infty$ find a further subsequence $(n_{k_s})_{s=1}^\infty$ such that

$$\mathbf{P}\left( m_{n_{k_s}} \left\| \mathbb{P}_{n_{k_s}}(\psi(\tilde{\epsilon}_{n_{k_s}}) \mathbf{1}_{\hat{\zeta}_{n_{k_s}} - \frac{(\cdot)}{m_{n_{k_s}}} < Z \leq \hat{\zeta}_{n_{k_s}}}) - (\cdot)\mathbb{P}(\psi(\epsilon))f(\zeta_0) \right\|_{K_s} > \frac{1}{s} \right) < \frac{1}{s^2}$$

$$\mathbf{P}\left( m_{n_{k_s}} \left\| \mathbb{P}_{n_{k_s}}(\psi(\tilde{\epsilon}_{n_{k_s}}) \mathbf{1}_{\hat{\zeta}_{n_{k_s}} < Z \leq \hat{\zeta}_{n_{k_s}} + \frac{(\cdot)}{m_{n_{k_s}}}}) - (\cdot)\mathbb{P}(\psi(\epsilon))f(\zeta_0) \right\|_{K_s} > \frac{1}{s} \right) < \frac{1}{s^2}.$$

The Borel-Cantelli Lemma will then imply that (VI) holds almost surely for the subsequence $(n_{k_s})_{s=1}^\infty$. Therefore, it suffices to show (63), (64), (65) and (66). First consider the case where $\psi(\cdot) = |\cdot|$ and a positive number $\eta > 0$. Let $t \in \mathbb{R}$ and write

$$r_n = n\mathbb{P}\left( e^{i\frac{m_n}{n}t|\epsilon|} - 1 - \frac{m_n}{n}t|\epsilon| \right) \mathbb{P}(\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}).$$

Then, $|r_n| \leq t^2 \sigma^2 \frac{m_n}{n} m_n \mathbb{P}(\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}) \to 0$. The characteristic function of $m_n \mathbb{P}_n(|\epsilon| \mathbf{1}_{\zeta_0 < Z \leq + \frac{\eta}{m_n}})$ can be written as

$$\varphi_n(t) = \left( 1 + i\frac{m_n}{n}t\mathbb{P}(|\epsilon|)\mathbb{P}(\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}) + \frac{r_n}{n} \right)^n \to e^{it\eta\mathbb{P}(|\epsilon|)f(\zeta_0)}$$

and therefore

$$m_n \mathbb{P}_n(|\epsilon| \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}) \xrightarrow{\mathbf{P}} \eta f(\zeta_0)\mathbb{P}(|\epsilon|).$$

But

$$\sup_{n \in \mathbb{N}} \left\{ \mathbf{E}\left( m_n \left\| \mathbb{P}_n(|\epsilon| \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}}) \right\|_{[0,\eta]} \right) \right\} < \infty$$

and hence the sequence of processes $\left( m_n \mathbb{P}_n(|\epsilon| \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}}) \right)_{n=1}^\infty$ is tight in $\mathcal{D}_{[0,\eta]}$. It follows that

$$m_n \mathbb{P}_n(|\epsilon| \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}}) \rightsquigarrow (\cdot) f(\zeta_0)\mathbb{P}(|\epsilon|) \quad \text{in} \quad \mathcal{D}_{[0,\eta]}$$

but since the limiting process is continuous and deterministic we actually obtain

$$\left\| m_n \mathbb{P}_n(|\epsilon| \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\cdot}{m_n}}) - (\cdot) f(\zeta_0)\mathbb{P}(|\epsilon|) \right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0. \qquad (69)$$

And with similar arguments one can also prove that

$$\left\| m_n \mathbb{P}_n(|\epsilon| \mathbf{1}_{\zeta_0 - \frac{(\cdot)}{m_n} < Z \leq \zeta_0}) - (\cdot) f(\zeta_0)\mathbb{P}(|\epsilon|) \right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0. \qquad (70)$$

Pick a positive number $\eta > 0$. Taking into account that $\epsilon \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}} = (y - \beta_0)\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}$ and the analogous result for $\tilde{\epsilon}_n$ with $\hat{\zeta}_n$ and $\hat{\beta}_n$ instead of $\zeta_0$ and $\beta_0$ we see that

$$m_n \left\| \mathbb{P}_n(|\tilde{\epsilon}_n|\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}}) - \mathbb{P}_n(|\epsilon|\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}}) \right\|_{[0,\eta]} \leq$$

$$m_n \left\| \mathbb{P}_n \left( |Y - \hat{\beta}_n| \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}} - \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right) \right\|_{[0,\eta]} +$$

$$m_n \left\| \mathbb{P}_n \left( \left( |Y - \hat{\beta}_n| - |Y - \beta_0| \right) \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right\|_{[0,\eta]}$$

and consequently

$$m_n \left\| \mathbb{P}_n(|\tilde{\epsilon}_n|\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}}) - \mathbb{P}_n(|\epsilon|\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}}) \right\|_{[0,\eta]} \leq$$

$$m_n \left\| \mathbb{P}_n \left( |Y - \beta_0| \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}} - \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right) \right\|_{[0,\eta]} +$$

$$|\hat{\beta}_n - \beta_0| m_n \left\| \mathbb{P}_n \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}} - \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right\|_{[0,\eta]} +$$

$$|\hat{\beta}_n - \beta_0| m_n \left\| \mathbb{P}_n \left( \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right\|_{[0,\eta]} . \tag{71}$$

We will show that each of the terms on the right-hand side of (71) goes to zero in probability. Since $n(\hat{\zeta}_n - \zeta_0) = O_{\mathbf{P}}(1)$, we know that for any $\delta > 0$ there is $R_\delta > 0$ such that $\mathbf{P}\left( n|\hat{\zeta}_n - \zeta_0| > R_\delta \right) < \delta$. Then,

$$\mathbf{P}\left( m_n \left\| \mathbb{P}_n \left( |Y - \beta_0| \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}} - \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right) \right\|_{[0,\eta]} > \delta \right) \leq \delta +$$

$$\mathbf{P}\left( m_n \mathbb{P}_n \left( |\epsilon|\mathbf{1}_{\zeta_0 - \frac{R_\delta}{n} < Z \leq \zeta_0} \right) > \frac{\delta}{3} \right) +$$

$$\mathbf{P}\left( m_n \left\| \mathbb{P}_n \left( |\epsilon|\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n} + \frac{R_\delta}{n}} \right) \right\|_{[0,\eta]} > \frac{\delta}{3} \right) +$$

$$\mathbf{P}\left( m_n |\alpha_0 - \beta_0| \mathbb{P}_n \left( \mathbf{1}_{\zeta_0 - \frac{R_\delta}{n} < Z \leq \zeta_0} \right) > \frac{\delta}{3} \right)$$

but from equations (69) and (70), and the fact that $\frac{m_n}{n} \to 0$, we actually get that all the terms of the right-hand side are asymptotically smaller than $\frac{\delta}{3}$. Thus,

$$\varlimsup_{n \to \infty} \mathbf{P}\left( m_n \left\| \mathbb{P}_n \left( |Y - \beta_0| \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}} - \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right) \right\|_{[0,\eta]} > \delta \right) < 2\delta. \tag{72}$$

An argument similar in spirit to the one just employed gives

$$\varlimsup_{n \to \infty} \mathbf{P}\left( m_n \left\| \mathbb{P}_n \left( \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}} - \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right) \right\|_{[0,\eta]} > \delta \right) < \delta \tag{73}$$

while equation (74), for $\xi = 0$, and the strong consistency of the least squares estimator give

$$|\hat{\beta}_n - \beta_0|m_n \left\|\mathbb{P}_n\left(\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}}\right)\right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0.$$

Then, combining the last identity with (71), (72) and (73) we get

$$\varlimsup_{\delta \to 0} \varlimsup_{n \to \infty} m_n \mathbf{P}\left(\left\|\mathbb{P}_n(|\tilde{\epsilon}_n|\mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}}) - \mathbb{P}_n(|\epsilon|\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}})\right\|_{[0,\eta]} > \delta\right) = 0.$$

Completely analogous arguments prove that

$$m_n\left\|\mathbb{P}_n(|\tilde{\epsilon}_n|\mathbf{1}_{\hat{\zeta}_n - \frac{(\cdot)}{m_n} < Z \leq \hat{\zeta}_n}) - \mathbb{P}_n(|\epsilon|\mathbf{1}_{\zeta_0 - \frac{(\cdot)}{m_n} < Z \leq \zeta_0})\right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0.$$

Since $\eta > 0$ was arbitrarily chosen, we have shown (IV) for $\psi(\cdot) = |\cdot|$. The case $\psi = |\cdot|^2$ is proven in a very similar manner. For the sake of brevity, we omit the proof.

Now, we consider the case where $\psi(x) = e^{i\xi x}$ for some $\xi \in \mathbb{R}$. Again, fix $\eta > 0$. We will proceed in the same way as before. Let $t \in \mathbb{R}$ and write

$$\rho_n = n\mathbb{P}\left(e^{i\frac{m_n}{n}t\cos(\xi\epsilon)} - 1 - \frac{m_n}{n}t\cos(\xi\epsilon)\right)\mathbb{P}(\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}).$$

Then, $|\rho_n| \leq t^2 \frac{m_n}{n} m_n \mathbb{P}(\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}) \to 0$. The characteristic function of $m_n \mathbb{P}_n(\cos(\xi\epsilon)\mathbf{1}_{\zeta_0 < Z \leq + \frac{\eta}{m_n}})$ can be written as

$$\varphi_n(t) = \left(1 + i\frac{m_n}{n}t\mathbb{P}(\cos(\xi\epsilon))\mathbb{P}(\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}) + \frac{r_n}{n}\right)^n \to e^{it\eta\mathbb{P}(\cos(\xi\epsilon))f(\zeta_0)}$$

and therefore

$$m_n\mathbb{P}_n(\cos(\xi\epsilon)\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}) \xrightarrow{\mathbf{P}} \eta f(\zeta_0)\mathbb{P}(\cos(\xi\epsilon)).$$

Applying the same arguments to the function $\sin(\xi\epsilon)$ we obtain that

$$m_n\mathbb{P}_n(\sin(\xi\epsilon)\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}) \xrightarrow{\mathbf{P}} \eta f(\zeta_0)\mathbb{P}(\sin(\xi\epsilon)).$$

and hence

$$m_n\mathbb{P}_n(e^{i\xi\epsilon}\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\eta}{m_n}}) \xrightarrow{\mathbf{P}} \eta f(\zeta_0)\varphi\xi = \eta f(\zeta_0)\mathbb{P}(e^{i\xi\epsilon}).$$

The same tightness argument that was applied to prove (69) can be used here to conclude that

$$\left\|m_n\mathbb{P}_n(e^{i\xi\epsilon}\mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{\cdot}{m_n}}) - (\cdot)f(\zeta_0)\mathbb{P}(e^{i\xi\epsilon})\right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0 \qquad (74)$$

and similarly

$$\left\| m_n \mathbb{P}_n(e^{i\xi\epsilon} \mathbf{1}_{\zeta_0 - \frac{(\cdot)}{m_n} < Z \leq \zeta_0}) - (\cdot) f(\zeta_0) \mathbb{P}(e^{i\xi\epsilon}) \right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0. \tag{75}$$

Using the triangular inequality together with the definition of $\tilde{\epsilon}_n$ we get

$$m_n \left\| \mathbb{P}_n(e^{i\xi\tilde{\epsilon}_n} \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}}) - \mathbb{P}_n(e^{i\xi\epsilon} \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}}) \right\|_{[0,\eta]} \leq$$

$$m_n \left\| \mathbb{P}_n \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}} - \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right\|_{[0,\eta]} +$$

$$m_n \left\| \mathbb{P}_n \left( (e^{i\xi(Y - \hat{\beta}_n)} - e^{i\xi(Y - \beta_0)}) \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right\|_{[0,\eta]}.$$

But (70) implies that

$$m_n \left\| \mathbb{P}_n \left( \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}} - \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0$$

while (74) applied when $\xi = 0$ and the strong consistency of $\hat{\beta}_n$ yield

$$m_n \left\| \mathbb{P}_n \left( (e^{i\xi(Y - \hat{\beta}_n)} - e^{i\xi(Y - \beta_0)}) \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right\|_{[0,\eta]} \leq$$

$$|\hat{\beta}_n - \beta_0| m_n \left\| \mathbb{P}_n \left( \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}} \right) \right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0.$$

Therefore,

$$m_n \left\| \mathbb{P}_n(e^{i\xi\tilde{\epsilon}_n} \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}}) - \mathbb{P}_n(e^{i\xi\epsilon} \mathbf{1}_{\zeta_0 < Z \leq \zeta_0 + \frac{(\cdot)}{m_n}}) \right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0$$

which together with (74) proves that

$$\left\| m_n \mathbb{P}_n(e^{i\xi\tilde{\epsilon}_n} \mathbf{1}_{\hat{\zeta}_n < Z \leq \hat{\zeta}_n + \frac{(\cdot)}{m_n}}) - (\cdot) f(\zeta_0) \mathbb{P}(e^{i\xi\epsilon}) \right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0.$$

With completely analogous arguments one shows

$$m_n \left\| \mathbb{P}_n(e^{i\xi\tilde{\epsilon}_n} \mathbf{1}_{\hat{\zeta}_n - \frac{(\cdot)}{m_n} < Z \leq \hat{\zeta}_n}) - (\cdot) f(\zeta_0) \mathbb{P}(e^{i\xi\epsilon}) \right\|_{[0,\eta]} \xrightarrow{\mathbf{P}} 0.$$

This proves that (VI) holds in probability.

We now proceed to prove that (V) and (VII) hold in probability. Before embarking in this task, we want to make the following remark. Consider that class of functions $\mathcal{C} := \{\epsilon \mathbf{1}_I(z) : I \subset \mathbb{R} \text{ is an interval}\}$. Then, this class has a square integrable envelope $|\epsilon|$ and $\mathbb{P}(\psi) = 0$ for any $\psi \in \mathcal{C}$. Therefore, the maximal inequality 3.1

from Kim and Pollard (1990) implies that $\|\mathbb{P}_n\|_{\mathcal{C}} = O_{\mathbf{P}}\left(n^{-\frac{1}{2}}\right)$. Similar observations also show that $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} = O_{\mathbf{P}}\left(n^{-\frac{1}{2}}\right)$. All these considerations, in addition with Corollary 3.1, (65), (66), (63) and (64) show that

$$\sqrt{m_n}(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathbf{P}} 0 \tag{76}$$

$$\sqrt{m_n}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathbf{P}} 0 \tag{77}$$

$$m_n(\hat{\zeta}_n - \zeta_0) \xrightarrow{\mathbf{P}} 0 \tag{78}$$

$$\sqrt{m_n}\|\mathbb{P}_n\|_{\mathcal{C}} \xrightarrow{\mathbf{P}} 0 \tag{79}$$

$$\sqrt{m_n}\left\|\mathbb{P}_n\left(|\epsilon|\mathbf{1}_{\zeta_0 - \frac{(\cdot)}{m_n} < Z \le \zeta_0 + \frac{(\cdot)}{m_n}}\right)\right\|_K \xrightarrow{\mathbf{P}} 0 \tag{80}$$

$$\sqrt{m_n}\left\|\mathbb{P}_n\left(|\tilde{\epsilon}_n|\mathbf{1}_{\hat{\zeta}_n - \frac{(\cdot)}{m_n} < Z \le \hat{\zeta}_n + \frac{(\cdot)}{m_n}}\right)\right\|_K \xrightarrow{\mathbf{P}} 0 \tag{81}$$

$$\sqrt{m_n}\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \xrightarrow{\mathbf{P}} 0 \tag{82}$$

for any compact set $K \subset \mathbb{R}$.

Let $\eta > 0$ be fixed. Take any subsequence $(n_k)_{k=1}^\infty$ and find a further subsequence $(n_{k_s})_{s=1}^\infty$ such that all the statements in the previous display happen almost surely with the compact set $K$ taken to be $K = [\zeta_0 - 2\eta, \zeta_0 + 2\eta]$. Now, for such a subsequence, there is $N \in \mathbb{N}$ such that $m_{n_{k_s}}|\zeta_0 - \hat{\zeta}_{n_{k_s}}| < \eta \; \forall \; s \ge N$. Then, for any $\delta > 0$ and $s \ge N$, the following inequalities are true

$$\sup_{|\hat{\zeta}_{n_{k_s}} - \zeta| < \delta^2} \left\{ |\mathbb{P}_{n_{k_s}}(\tilde{\epsilon}_{n_{k_s}} \mathbf{1}_{\zeta \wedge \hat{\zeta}_{n_{k_s}} < Z \le \zeta \vee \hat{\zeta}_{n_{k_s}}})| \right\} \le |\hat{\alpha}_{n_{k_s}} - \alpha_0| + |\hat{\beta}_{n_{k_s}} - \beta_0| +$$

$$\mathbb{P}_{n_{k_s}}\left(|\tilde{\epsilon}_{n_{k_s}}|\mathbf{1}_{\hat{\zeta}_{n_{k_s}} - \frac{\eta}{m_{n_{k_s}}} < Z \le \hat{\zeta}_{n_{k_s}} + \frac{\eta}{m_{n_{k_s}}}}\right) + \mathbb{P}_{n_{k_s}}\left(|\epsilon|\mathbf{1}_{\zeta_0 - \frac{2\eta}{m_{n_{k_s}}} < Z \le \zeta_0 + \frac{2\eta}{m_{n_{k_s}}}}\right) + \left\|\mathbb{P}_{n_{k_s}}\right\|_{\mathcal{C}}$$

$$\sup_{|\hat{\zeta}_{n_{k_s}} - \zeta| < \delta^2} \left\{ |\mathbb{P}_{n_{k_s}}(\tilde{\epsilon}_{n_{k_s}} \mathbf{1}_{Z \le \zeta \wedge \hat{\zeta}_{n_{k_s}}})| + |\mathbb{P}_{n_{k_s}}(\tilde{\epsilon}_{n_{k_s}} \mathbf{1}_{Z > \zeta \vee \hat{\zeta}_{n_{k_s}}})| \right\} \le |\hat{\alpha}_{n_{k_s}} - \alpha_0| + |\hat{\beta}_{n_{k_s}} - \beta_0| +$$

$$\mathbb{P}_{n_{k_s}}\left(|\tilde{\epsilon}_{n_{k_s}}|\mathbf{1}_{\hat{\zeta}_{n_{k_s}} - \frac{\eta}{m_{n_{k_s}}} < Z \le \hat{\zeta}_{n_{k_s}} + \frac{\eta}{m_{n_{k_s}}}}\right) + \mathbb{P}_{n_{k_s}}\left(|\epsilon|\mathbf{1}_{\zeta_0 - \frac{2\eta}{m_{n_{k_s}}} < Z \le \zeta_0 + \frac{2\eta}{m_{n_{k_s}}}}\right) + \left\|\mathbb{P}_{n_{k_s}}\right\|_{\mathcal{C}}.$$

These last inequalities together with (76)-(82) imply that

$$\varlimsup_{s \to \infty} \sqrt{m_{n_{k_s}}} \sup_{|\hat{\zeta}_{n_{k_s}} - \zeta| < \delta^2} \left\{ |\mathbb{P}_{n_{k_s}}(\tilde{\epsilon}_{n_{k_s}} \mathbf{1}_{\zeta \wedge \hat{\zeta}_{n_{k_s}} < Z \le \zeta \vee \hat{\zeta}_{n_{k_s}}})| \right\} = 0 \; a.s.$$

$$\varlimsup_{s \to \infty} \sqrt{m_{n_{k_s}}} \sup_{|\hat{\zeta}_{n_{k_s}} - \zeta| < \delta^2} \left\{ |\mathbb{P}_{n_{k_s}}(\tilde{\epsilon}_{n_{k_s}} \mathbf{1}_{Z \le \zeta \wedge \hat{\zeta}_{n_{k_s}}})| + |\mathbb{P}_{n_{k_s}}(\tilde{\epsilon}_{n_{k_s}} \mathbf{1}_{Z > \zeta \vee \hat{\zeta}_{n_{k_s}}})| \right\} = 0 \; a.s.$$

The previous equations show that (6) and (7) in (V) as well as (VII) hold with probability one for the subsequence $(n_{k_s})_{s=1}^\infty$. We conclude by noting that if $\kappa = $

$\inf\limits_{z\in[a,b]}\{f(z)\}$, then the mean value theorem implies

$$\inf_{\frac{1}{\sqrt{m_{n_{k_s}}}}\leq|\zeta-\hat{\zeta}_{n_{k_s}}|<\delta^2}\left\{\frac{1}{|\zeta-\zeta_{n_{k_s}}|}\mathbb{P}_{n_{k_s}}\left(\mathbf{1}_{\zeta\wedge\hat{\zeta}_{n_{k_s}}<Z\leq\zeta\vee\hat{\zeta}_{n_{k_s}}}\right)\right\}\geq\kappa-\sqrt{m_{n_{k_s}}}\left\|\mathbb{P}_{n_{k_s}}-\mathbb{P}\right\|_{\mathcal{F}}$$

which in consequence shows

$$\varliminf_{s\to\infty}\inf_{\frac{1}{\sqrt{m_{n_{k_s}}}}\leq|\zeta-\hat{\zeta}_{n_{k_s}}|<\delta^2}\left\{\frac{1}{|\zeta-\zeta_{n_{k_s}}|}\mathbb{P}_{n_{k_s}}\left(\mathbf{1}_{\zeta\wedge\hat{\zeta}_{n_{k_s}}<Z\leq\zeta\vee\hat{\zeta}_{n_{k_s}}}\right)\right\}\geq\kappa>0 \quad \text{a. s.}$$

This finishes the proof. $\qquad\square$

# References

Abrevaya, J. and Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica*, 73:11751204.

Bickel, P. J., Götze, F., and van Zwet, W. R. (1997). Resampling fewer than $n$ observations: gains, losses, and remedies for losses. *Statist. Sinica*, 7:1–31.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York, NY, USA.

Billingsley, P. (1986). *Probability and Measure.* John Wiley, New York, NY, USA.

Bose, A. and Chatterjee, S. (2001). Generalised bootstrap in non-regular m-estimation problems. *Stat. Prob. Lett.*, 55:319–328.

Dempfle, A. and Stute, W. (2002). Nonparametric estimation of a discontinuity in regression. *Statist. Neerlandica*, 56:233–242.

Dudley, R. M. (2002). *Real Analysis and Probability.* Cambridge University Press, Cambridge, UK.

Dümbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.*, 19:1471–1495.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38. SIAM, Philadelphia, USA.

Ethier, S. and Kurtz, T. (2005). *Markov Processes, Characterization and Convergence.* John Wiley & Sons, New York, NY, USA.

Freedman, D. (1981). Bootstrapping regression models. *Ann. Statist.*, 9:1218–1228.

Gijbels, I., Hall, P., and Kneip, A. (1999). On the estimation of jump point in smooth curves. *Ann. Inst. Statist. Math.*, 51:231–251.

Gijbels, I., Hall, P., and Kneip, A. (2004). Interval and band estimation for curves with jumps. *J. Appl. Probab.*, 41:65–79.

Hazma, K. (1995). The smallest uniform upper bound on the distance between the mean and median of binomial and poisson distributions. *Statist. Probab. Lett.*, 23:21–25.

Hŭsková, M. and Kirch, C. (2008). Bootstrapping confidence intervals for the change-point of time series. *J. Time Ser. Anal.*, 29:947–972.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statis.*, 18:191–219.

Kosorok, M. (2008a). Bootstrapping the grenander estimator. *Beyond parametrics in interdisciplinary research: Festschrift in honor of professor Pranab K. Sen, IMS Collections*, 1:282–292.

Kosorok, M. (2008b). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York, NY, USA.

Kosorok, M. and Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.*, 35:957–989.

Koul, H., Qian, L., and Surgailis, D. (2003). Asymptotics of m-estimators in two-phase linear regression models. *Stoc. Proc. Applic.*, 103:123–154.

Lan, Y., Banerjee, M., and Michailidis, G. (2009). Change-point estimation under adaptive sampling. *Ann. Statist.*, 37:1752–1791.

Lee, S. and Pun, M. (1981). On $m$ out of $n$ bootstrapping for nonstandard m-estimation with nuisance parameters. *J. Amer. Statist. Assoc.*, 9:1185–1197.

Möhle, M. (2005). Convergence results for compound poisson distributions and applications to the standard luria-delbrck distribution. *C. R. Acad. Sci. Paris*, 42:620–631.

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.*, 42:1285–1295.

Pons, O. (2003). Estimation in a cox regression model with a change-point according to a threshold covariate. *Ann. Statist.*, 31:442–463.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, New York, NY.

Sen, B., Banerjee, M., and Woodroofe, M. (2010). Inconsistency of bootstrap: The grenander estimator. 38:1953–1977.

Silverman, B. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, 6:177–184.

Van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York, NY, USA.

Williams, D. (1991). *Probability with Martingales.* Cambridge University Press, Cambridge, UK.

Wu, C. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.*, 14:1261–1295.